

***Design Document for
12th Grade NAEP Preparedness Research
Judgmental Standard Setting Studies***

**Setting Standards on the National Assessment of
Educational Progress in Reading and Mathematics
for 12th Grade Preparedness**

National Assessment Governing Board

July 2010

Design Document for 12th Grade NAEP Preparedness Research Judgmental Standard Setting Studies

**Setting Standards on the National Assessment of
Educational Progress in Reading and Mathematics
for 12th Grade Student Preparedness**

A draft of this report was prepared by ACT, Inc., under contract ED-06-CO-0098 with the National Assessment Governing Board.

Design Document for 12th Grade NAEP Preparedness Research Judgmental Standard Setting

Table of Contents

Introduction.....	1
BACKGROUND FOR THE DEVELOPMENT OF A DESIGN DOCUMENT.....	1
PURPOSE OF THIS DOCUMENT	1
ORGANIZATION OF THIS DOCUMENT	1
 Standard-setting Process	2
DESIGN	2
<i>Standard-setting Workshops</i>	2
<i>Pilot Study</i>	2
<i>Timing and Logistics</i>	2
<i>Replicate Panels and Composition of Panels</i>	3
<i>Item Pool Division</i>	3
<i>Standard-setting Methodology</i>	4
<i>Facilitators</i>	4
<i>Room Setup</i>	4
<i>Document Preparation for the Meeting</i>	5
<i>Logistical Support for the Studies</i>	5
IDENTIFICATION OF THE PANELISTS	6
<i>Qualifications of Standard-setting Panelists for Higher Education</i>	6
Mathematics Panelists	6
Reading Panelists	7
Panelist Selection.....	8
<i>Standard-setting Panels for Occupational Training Programs</i>	9
Identifying Occupational Training Programs.....	9
The Panelists	9
Panelist Selection.....	10
BORDERLINE STUDENT DESCRIPTIONS	10
BRIEFING MATERIALS FOR WORKSHOPS.....	11
STANDARD-SETTING WORKSHOP	12
<i>Orientation</i>	12
General Orientation.....	13
Orientation to the Modified bookmark Method and Materials	13

<i>Round 1: Understanding the Assessment and Defining the Borderline</i>	
Student	20
Overview of Round 1	20
Taking NAEP	20
The Assessment Framework.....	21
Review and Revision of the Borderline Student Description.....	21
Item Review.....	21
Borderline Student Description	25
Placing the Bookmark.....	25
<i>Round 2: Understanding Student Performance</i>	27
Overview of Round 2	27
Feedback from Round 1.....	27
Whole Booklet Feedback	30
Review and Revise the Borderline Student Description	35
Round 2 Cut Score Recommendations	35
<i>Round 3: Impact Data</i>	36
Overview of Round 3	36
Feedback from Round 2.....	36
Impact Data and Discussion.....	36
Round 3 Cut Score Recommendations	37
<i>Post-Round 3 Activities</i>	37
Feedback from Round 3.....	37
Impact Data Questionnaire	37
Selection of Exemplar Items.....	39
<i>Process Evaluation</i>	39
Information Processing	39
Validity Evidence.....	41
<i>Procedural Validity</i>	42
<i>Internal Validity</i>	43
Reporting.....	45
References.....	46

Appendices

Appendix A:	Sample Agenda for a Standard-setting Meeting.....	A1-A5
Appendix B:	Definition of 12 th Grade Student Preparedness Resolution, March 2009	B1
Appendix C:	Comparative Data Questionnaire	C1-C2
Appendix D:	Evaluation Questionnaires	D1-D22
Appendix E:	List of Potential Exemplar Occupations.....	E1

List of Tables

Table 1:	Types of validity.....	41
Table 2:	General evaluation questions used at the completion of each meeting	42

List of Figures

Figure 1:	Room and table setup	5
Figure 2:	A simplified item map as spatially representative of achievement from low to high	14
Figure 3:	The relationship of the RP criterion to an item's scale value.....	15
Figure 4:	Illustration of how items are displayed on an item map.....	16
Figure 5:	Illustration of how items are ordered by difficulty in the Ordered Item Book (OIB). ..	17
Figure 6:	Illustration showing item location and information location for Item C10 in the OIB	18
Figure 7:	Illustration of the information on an OIB page	19
Figure 8:	Illustration of round 1 bookmark placement	20
Figure 9:	Illustration showing information location for Item C12 in the CROIB.....	23
Figure 10:	Illustration showing the location of extended CR item score levels in the OIB.....	24
Figure 11:	Illustration of the relationship between bookmark placement and the “mastery” of items at and below the cut score using 0.67 as the RP criterion.....	26
Figure 12:	Illustration of the range of uncertainty in a bookmark placement	27
Figure 13:	Illustration of cut score distribution chart showing the distribution of panelists' cut scores	28

Figure 14: Illustration of Primary Item Map showing the round 1 panel cut score and the location of a panelist's bookmarked item	29
Figure 15: Illustration of the comparison of the panel cut score and a panelist's bookmarked item in the Ordered Item Book	30
Figure 16: Illustration of Booklet Score Plot for a form, showing the round 1 cut score and the score of each student booklet on the achievement scale.....	31
Figure 17: Sample Booklet Score Chart for group A showing the median, high and low cut scores, and the location of a panelist's round 1 cut score	32
Figure 18: Illustration of an Item Score Table for Form C, with the items listed from hardest to easiest, top to bottom, in the left most column and the student booklets listed from lowest to highest score, left to right, in the top row with the number correct score on the booklet provided below the booklet identifier.....	34
Figure 19: Illustration of a Cut Score Proportion Chart identifying the percent of students scoring at or above every fifth score level.....	38

Introduction

BACKGROUND FOR THE DEVELOPMENT OF A DESIGN DOCUMENT

The National Assessment Governing Board adopted a Program of Preparedness Research in March 2009. Several categories of research studies were recommended to produce results for reporting 12th grade preparedness for the 2009 grade 12 National Assessment of Educational Progress (NAEP) in reading and mathematics. The categories included content alignment studies, statistical relationship studies, judgmental standard-setting studies, and surveys.

The Governing Board will conduct a series of judgmental standard-setting studies to produce preparedness reference points on the NAEP scale for entry into job training programs and for placement in college credit-bearing courses. These preparedness reference points will represent the academic knowledge and skills required for postsecondary course and training placement. In order to maximize the standardization of judgmental standard-setting studies within and across postsecondary areas, the Governing Board developed this design document to be used for all judgmental standard-setting studies of preparedness for NAEP.

PURPOSE OF THIS DOCUMENT

The purpose of this design document is to describe the procedures for the conduct of judgmental standard setting on the 2009 NAEP for grade 12 reading and mathematics. Specifically, the standard-setting activity is described in relation to the reading and mathematics skills and knowledge needed to qualify (a) for placement in entry-level credit-bearing postsecondary courses or (b) for placement in training programs in each of five occupations to be determined by the Governing Board. The design document describes the process to be implemented and the types of staff required to do the work. A modified bookmark method will be used as the judgmental standard-setting methodology. The goal is to maximize the comparability across the judgmental studies within and across the postsecondary activities. It is likely that more than one vendor will be required to implement the complete set of studies planned, and the information in this design document will allow the Governing Board to contract with multiple vendors to conduct comparable standard-setting activities for higher education and multiple occupational training programs.

ORGANIZATION OF THIS DOCUMENT

The design document is organized into three major sections.

The **Standard-setting Process** section describes the methodology for conducting the standard setting. This includes:

- higher education versus occupational training studies
- methodology, meeting format, facilitators, room setup, and logistical support
- identification of panelists
- briefing materials
- pilot study
- specific details about how the standard-setting meeting will be conducted

The second section, **Information Processing**, describes the information processing requirements to conduct the standard setting. The **Validity Evidence** section describes methods for obtaining procedural and internal validity of the standard-setting process.

Standard-setting Process

The standard setting process includes, but is not limited to, recruiting panelists, developing the descriptions of borderline preparedness performance, preparing materials and training protocols for the standard-setting meetings, conducting a pilot study, and conducting the operational standard-setting study. These steps are the same whether the standard setting is for higher education or for occupational training programs. Details of the procedures may differ, however. The work in the standard-setting process is described below generically. When there are differences, as in the selection criteria for panelists, specific directions are provided for each type of postsecondary program. Topics covered in the standard setting process are design, identification of the panelists, briefing materials, and the standard setting meeting.

DESIGN

Standard-setting Workshops

For purposes of this design document, a standard-setting workshop consists of cut-score studies for both mathematics and reading in higher education or for one occupation. Workshops for multiple occupations may be conducted at the same time if resources are available. The methodology for the pilot study and the operational standard-setting workshop is the same for both higher education and workforce training programs. A modified bookmark method will be used as the judgmental standard-setting methodology.

Pilot Study

The pilot study is to be implemented using exactly the same procedures planned for the operational standard-setting meeting in order to determine whether modifications to training, instructions, materials, timing, and so forth are needed. For this reason the detailed description of the design decisions and of how the standard-setting will be conducted described below applies both to the pilot study and to the operational standard-setting meeting. The Governing Board plans to conduct standard-setting studies for both the mathematics and reading NAEP of job training preparedness for approximately five different occupations and a pilot study of both the 12th grade mathematics and reading for one occupation. If multiple vendors are selected to conduct judgmental standard-setting studies for different occupations, then instead of having one pilot study for one occupation, there will be one pilot for each vendor. For example, if one vendor conducts judgmental standard-setting studies for two occupations and another vendor conducts judgmental standard-setting studies for three other occupations, each vendor will have the opportunity to conduct its own pilot study. For college course placement, a pilot study of both mathematics and reading will be implemented in preparation for the operational studies for each subject.

Timing and Logistics

The standard-setting meeting is planned to last at least 3½ days. A sample agenda for a standard-setting meeting is shown in Appendix A.

The meeting will require one large meeting room for the kick off meetings for each workshop, one room for use by staff for preparation of meeting materials throughout the meeting, and four

breakout rooms—two for mathematics standard setting and two for reading standard setting. The standard-setting workshop will start with an orientation session for all panelists. The panelists will then take a form of the assessment in the subject for which they will be setting a preparedness cut score. Following the general training in the method and materials, process and content (subject matter) facilitators will lead panelists through the specific tasks for each round of the modified bookmark method.

Replicate Panels and Composition of Panels

A replicate panel design will be used for these studies. Replication panels are included in the study design as a way of estimating how consistently the cut score is set across panels—to assess the reliability of the judgments. The two replicate panels for each subject will work independently. The exception to this is for general orientation and training, and development of the borderline performance descriptions for preparedness in each subject.

As one step to avoid “contamination” of the standard-setting process (influence by replicate panel A on the cut score decision of replicate panel B), different scales will be used in the process. Each scale will be a linear transformation of the NAEP reporting scale. A different NAEP-like scale is to be used for each panel/cut score study for both higher education and occupational training programs, and a different NAEP-like scale will be used for the replicate panels within a subject. Further, separate process facilitators will be assigned to replicate panels, and replicate panels will meet in different rooms so that the discussions will not be overheard.

One operational workshop will include 40 panelists. Twenty panelists will be required for each operational standard-setting study in each subject (replicate panels of 10 each), and 12 panelists will be required for each pilot study (6 panelists for each replicate panel). For each subject (mathematics or reading) panelists will be assigned to one of two replicate panels: panel A or panel B. Each replicate panel will be further divided into table groups for individual work and to facilitate group discussion. The demographic attributes of panelists will be considered when assigning members to replicate panels and tables, and the panelists will be selected to maximize equivalence of the replicate panels. Similarly, table group assignments will be made to maximize equivalence across the groups. Otherwise, the assignments shall be random. The goal is to have replicate panels and tables as equal as possible with respect to panelist type (i.e. educator role), gender, geographic region, and race/ethnicity.

Panelists for the judgmental standard setting studies for occupational training programs will be selected from instructors of those programs. Panelists for higher education will be selected from educators who prepare students for entry-level credit-bearing course work (e.g., high school teachers or postsecondary instructors of remedial classes) and college faculty teaching the entry-level credit-bearing courses in the subject. (Panelist recruitment will be discussed in greater detail in a later section.) In all cases, the most highly qualified and outstanding candidates should be selected to serve as panelists for the studies.

Item Pool Division

All items in the 2009 NAEP item pool will be made available to the contractor, for use in the study for each subject—reading and mathematics. These items are secure, and the security of the items must be maintained at all times. To reduce the number of items for each panelist to judge, the contractor will divide the item pool for each assessment into equivalent, but overlapping, rating pools for replicate panels A and B. Equivalence will be evaluated with regard to the following criteria: (a) mean and variability of item difficulty, (b) representation of critical

framework variables, and (c) percent of items of each type. To the extent possible, the equivalence criteria will be met by assigning blocks of items to the pools. One block will be common to the two replicate (rater) panels. For NAEP, a block of items is a set of items administered over a 25-minute period, and each block is paired with one other block to form an assessment booklet—pairings occur in various sequences as part of the design of the assessment.

Standard-setting Methodology

A modified bookmark method will be used as the judgmental standard-setting methodology. This type of methodology has been used in recent NAEP achievement levels-setting work, starting in 2005. (See, for example, the process report for 2006 grade 12 economics at www.nagb.org/publications/2006-g12th-econ-process-report.pdf.) A response probability of .67 has been used in previous NAEP achievement levels-setting studies, and response probability (RP) .67 shall be used for this study. This value is used for all items, both constructed response and multiple choice, and there is no correction for guessing. Key materials required for this methodology include an item map and an Ordered Item Book (OIB). Both of these materials use a common score scale based on the psychometric model used to scale the items. This score scale simultaneously represents both the difficulty of items on the assessment and the performance of students on these items. The scale values by which items are located on the item maps used in the modified bookmark process will be a linear transformation of the composite scale that is used for reporting assessment results in official NAEP reports. Likewise, item sequence in the OIB will be based on the item's location on the composite scale.

The composite scale will be used for item mapping because NAEP results are reported on the composite scale. The composite scale for grade 12 NAEP is constructed from an item response theory (IRT) calibration of items for each subscale, and it will be necessary to regress the subscale item characteristic curve, or response probability (RP), onto the composite scale. The methodology used for NAEP achievement levels-setting has been based on the regression methods used by Schulz, Lee, and Mullen (2005) in a study with grade 8 mathematics NAEP items.

Facilitators

Four process facilitators and two content facilitators will be needed for each standard-setting workshop both for college course placement and for each occupation selected for NAEP reporting. A standard-setting workshop includes both mathematics and reading; two process facilitators (one for each replicate panel) and one content facilitator will be needed for each subject. All facilitators must be excellent presenters, have extensive experience in facilitation of similar processes, and at least some prior experience with the bookmark standard-setting method. Content facilitators should be selected from among the members of the 2009 framework development panels for grade 12 NAEP mathematics or reading.

Room Setup

Because the workshop will involve multiple cut score studies, and multiple sets of facilitators, and multiple groups of panelists, multiple separate medium-sized meeting rooms will be required for the standard-setting workshops. Figure 1 illustrates a typical room setup for each replicate panel for either mathematics or reading. For each subject, one of the replicate panel rooms should be large enough to accommodate all panelists for a single subject/occupation in order to discuss issues common to the two replicate panels in each subject, such as to refine the borderline performance descriptions that must be common to both replicate panels.

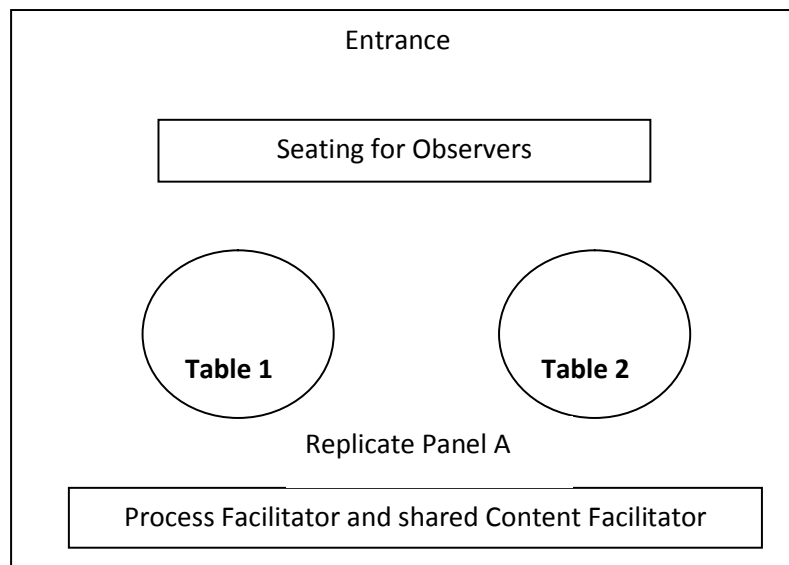


Figure 1: Room and table setup

The contracting officer's representative (COR) for the National Assessment Governing Board, members of the Board's Committee for Standards, Design and Methodology, and the Governing Board's technical advisors may attend the standard setting meetings as observers.

Document Preparation for the Meeting

Holistic feedback has been regularly provided in NAEP achievement levels-setting procedures since 1993, and it is to be included in these studies. NAEP test booklets used by students in the 2009 grade 12 assessment will be used for providing holistic feedback in the process. The Governing Board's COR will arrange to make these booklets available to the contractor for this purpose. Based on recommendations from the science ALS study (ACT, 2010), student booklets will be selected before the standard-setting meeting to avoid labor and time-intensive on-site materials preparation. Booklets will be pre-selected at a fixed interval along the scale (e.g., every 15 points). The determination of the range for the booklets and the interval between the booklets will be based on data from the 2009 assessment and in consultation with the Governing Board's COR. Booklets will only be selected for a portion of the entire scale score range—not at the highest or lowest score ranges.

All meeting materials and presentations must be developed in time for review and feedback by Governing Board staff prior to use in the standard-setting meetings.

Logistical Support for the Studies

A critical aspect of meeting preparations is the need for sophisticated logistical support. Meeting materials must be produced, shipped, and stored in a secure and accurate manner. Logistics for each standard-setting workshop include coordinating meeting rooms, lunches and breaks, audio-visual equipment and resources, computers, printers, and copier equipment; coordinating communications with panelists; arranging or coordinating reservations for travel and lodging accommodations; obtaining confidentiality agreements from all participants; managing travel

reimbursement; and assuring security and quality control over all materials and procedures. In addition, psychometric and technical assistance are needed for each study.

The selection of meeting site requires a site visit and inspection by an experienced logistic planner familiar with the requirements of this type of panel meeting. The meeting location must be accessible to panelists from across the nation. The locale must provide the opportunity for dining and entertainment that is easily accessible and safe for panelists. Pre-inspection of the site avoids surprises and will be useful in guiding meeting participants on issues such as transportation from the airport to the hotel. The inspection is essential to assure the suitability of the meeting and breakout rooms individually and relative to one another and to the “work room” where data processing will take place and materials collected and stored.

All services that may be needed for the studies must be available at the selected site. Experience has shown that prompt repair service is as important as the availability of the equipment.

The vendor is encouraged to have panelists use computers in the process, to the extent that this increases efficiency and effectiveness of the process. To date, computers have not been incorporated into the NAEP achievement levels-setting process, but the Governing Board encourages their use.

IDENTIFICATION OF THE PANELISTS

It is essential that the panelists be carefully selected and broadly representative. In order to encourage panelist participation, panelists will be given a \$500 honorarium for participation in the study. The primary requirement for selection is that the panelists know the reading or mathematics requirements for placement—without remediation—into higher education courses or for occupational training programs. In addition, both demographic characteristics and panel size are key considerations in the selection of panelists. Procedures for selecting panelists for higher education and occupational training programs are described separately below.

Qualifications of Standard-setting Panelists for Higher Education

The panelists who are recruited to set mathematics or reading standards for entry-level credit-bearing college courses need to have a good conceptual understanding of the reading or mathematics knowledge and skills needed to qualify for placement in the appropriate course. It is expected that the qualifications for reading and mathematics panelists will differ. While most higher education courses require the ability to read, few instructors in the disciplines (e.g., social sciences) have training in reading. On the other hand, instructors of higher education courses that require mathematics typically have training in mathematics.

Mathematics Panelists

Mathematics standard-setting panelists need to be informed and knowledgeable about the mathematics requirements for placement, without remediation, in an entry-level, credit-bearing course in mathematics. The course must be of the type to fulfill a general education course requirement in mathematics and be offered in the mathematics department/program, but the course would not necessarily fulfill a requirement for a major in mathematics, engineering, or pre-medicine. Individuals with the following types of experience will be appropriate candidates for the mathematics panels:

- Instructors of 2- and 4-year higher education entry-level credit-bearing mathematics courses that fulfill general education requirements for a four-year degree program
- Instructors of remedial/developmental mathematics courses, such as intermediate algebra, in postsecondary institutions
- Postsecondary mathematics instructors who have participated directly in development of entry-level mathematics placement tests for a postsecondary institution
- Postsecondary mathematics instructors of entry-level mathematics courses who have participated directly in development of high-school-to-college transition projects
- Grade 12 high school mathematics instructors who have worked with developers of college admission or placement tests or who have worked on high-school-to-college transition projects

In addition, panelists should have the following qualifications:

- At least five years of grade 12 or postsecondary mathematics teaching experience in courses appropriate to the targeted entry-level courses for student placement. For high school teachers, this may include teaching courses that count for college credit or teaching in dual enrollment programs.
- Judged to be very good in their professional performance by a supervisor or someone in the position to make that judgment.

Research may be needed to determine the most effective way to identify and draw samples of, such individuals for recruitment purposes. The list of recommended panelists and their qualifications must be presented for approval by the Governing Board before recruitment of panelists begins.

Reading Panelists

Reading standard-setting panelists need to be informed and knowledgeable about the reading requirements for course placement, without remediation, in courses that have an intensive or extensive reading demand and fulfill a general education requirement. The course may be in literature or in one of the social sciences. Some courses in the humanities may also be appropriate for this purpose. Individuals with the following types of experience will be appropriate candidates for the reading panels:

- Instructors of 2- and 4-year higher education entry-level credit-bearing English language arts courses that fulfill general education requirements for a four-year degree program
- Instructors of remedial/developmental reading courses in postsecondary institutions
- Postsecondary instructors who specialize in reading instruction or curriculum
- Postsecondary English language arts instructors who have participated directly in development of entry-level reading/English language arts placement tests for a postsecondary institution
- Postsecondary English language arts instructors of entry-level courses who have participated directly in development of high-school-to-college transition projects
- Grade 12 high school English language arts instructors who have participated directly in development of college placement tests or who have worked with developers of college admission or placement tests or who have worked on high school-to-college transition projects

In addition, panelists should have the following qualifications:

- At least five years of grade 12 or postsecondary English language arts teaching experience in courses appropriate to the targeted entry-level courses for student placement. For high school teachers, this may include teaching courses that count for college credit or teaching in dual enrollment programs.
- Judged to be very good in their professional performance by a supervisor or someone in the position to make that judgment.

Research may be needed to determine the most effective way to identify and draw samples of, such individuals for recruitment purposes. The list of recommended panelists and their qualifications must be presented for approval by the Governing Board before recruitment of panelists begins.

Panelist Selection

Once the qualifications required for mathematics and reading panels are finalized, a method for identifying and sampling members of well-defined target populations will be used to select panelists. By using aspects of sampling methodology, it is possible to select broadly representative panels through which diverse points of view can be expressed, and it will be possible to replicate the selection process.

Panelists will be selected for the pilot study and the standard-setting workshops based on their qualifications, the need to equalize the replicate panels, and the availability of panelists to participate. Every effort must be made to meet the targeted proportions of persons with the desired attributes.

Highest priority is given to the most qualified nominees within each target population. A scoring scheme should be developed to rate the qualifications of candidates, and it may differ slightly for reading and mathematics. The qualifying credentials of nominees will be evaluated and scored based on the number and importance of the credentials presented. Persons for whom little or no information is provided and persons having no distinguishing credentials will score low. Persons having extensive qualifying credentials will score very high.

Persons with the highest scores will be given highest priority for selection by placing the best qualified candidates at the beginning of the candidate list. The selection process will then select from the list of most highly qualified candidates to be representative of the following attributes:

- Gender
- Race/ethnicity
- Location (geographical region and rural/urban setting)
- Type of mathematics or reading experience
- Type of institutional affiliation (e.g., high school, community college, open-admission 4-year public institution, private 4-year college)

Considering the small panel size, it will not be possible to ensure that each panel is representative with respect to each combination of characteristics (e.g., Hispanic females in private 4-year colleges in the central U.S.). Further, because more males than females are engaged in college-level instruction, and more females than males are engaged in English language arts instruction, gender equivalence is not possible across panels. The primary goal of this plan is to produce broadly representative panels of persons well-qualified to make the judgments required by the standard-setting process. The list of recommended panelists and

their qualifications must be presented for approval by the Governing Board before recruitment of panelists begins.

Standard-setting Panels for Occupational Training Programs

It is recommended that standard-setting panelists for occupational training programs be recruited from instructors of such programs. Each standard setting workshop will include replicate panels for reading and for mathematics for one job training program. Multiple workshops may be conducted simultaneously for more than one job training program, but this is not a requirement. The availability of resources and the ability to demonstrate effective quality control procedures will determine the number of simultaneous studies conducted at a site.

Panels of instructors who know the mathematics or reading requirements for placement in training program in the occupational area of the workshop must be recruited for the studies. The panelists may be recruited from occupational training settings that represent the range of training environments for each targeted occupation. Once the occupations for study are identified, research will be conducted to determine the types of training environments for each occupation (e.g., community college programs, trade schools).

Identifying Occupational Training Programs

The Governing Board is in the process of identifying five occupations for these studies. The occupations for these studies will be selected by the Governing Board from a list of 20 occupations listed in Appendix E. The types of occupations to be selected for these studies are described on pages 19-22 of ***Technical Panel on Preparedness Research: Final Report*** located at (www.nagb.org/publications/PreparednessFinalReport.pdf). For each occupation, there may be many colleges or training programs from which to recruit panelists, so a method of sampling the programs is to be used. A list of training programs for the occupation will be developed and a sample will be drawn of these programs. Each of the selected programs will be contacted and asked to nominate one or two instructors/job trainers for that occupation as candidates to be on the mathematics or reading NAEP panels. The candidates should know the subject matter (reading or mathematics) required for entry into the job training program. Stratified sampling procedures should be used for selecting the sample of schools/occupational training programs to nominate candidates. The training programs from which panelists will be selected should be sampled proportionally to assure broadly representative panels according to types of programs, geographical region and location, and other attributes judged important to the options available for entry-level training in that occupation.

Occupational training programs should be accredited by an appropriate agency or organization. For programs in higher education institutions, the Higher Education Directory can be used to identify appropriate programs.

The Panelists

Occupational training programs panelists must be well informed and knowledgeable about the occupational and entry-level training requirements for students. In particular, the panelists nominated for each subject panel must know the subject-specific requirements needed for entry in the job training program. While qualifications may vary somewhat by occupation, they will always include the following:

- At least five years of experience teaching entry-level courses in the occupational training program

- Judged to excellent in their professional performance by a supervisor or someone in the position to make that judgment.

The list of recommended panelists and their qualifications must be presented for approval by the Governing Board before recruitment of panelists begins.

Panelist Selection

Please see the section on panelist selection for higher education. The same requirements and considerations hold for selection of panelists for the job training judgmental standard setting workshops. While the proportion of males and females that serve as trainers/instructors will vary by subject and by occupation, the proportion of males and females selected for panels should be proportional to the numbers in the occupational training programs. Similarly, to the extent possible, the institutional affiliation of panelists should be representative of the types of institutions offering training programs in the occupation.

BORDERLINE STUDENT DESCRIPTIONS

The development of the borderline performance descriptions will be informed by the expertise and experience of the panelists and will be based on the content assessed by NAEP for each subject. To the extent that knowledge and skills deemed necessary and sufficient for borderline performance are not represented in the NAEP for the subject, this shall be noted and the knowledge and skills *not* included in NAEP shall be clearly documented. An initial element of the borderline description development process should include a determination by the panelists that there is a sufficient match of borderline requirements and NAEP content to proceed with the standard-setting workshop. The description of borderline performance is extremely important to the standard-setting study because this description is used as the standard for judging where to place the bookmark to represent preparedness for each postsecondary activity. Borderline preparedness performance descriptions will be developed by panelists for each subject, with the assistance of the content facilitator, in four stages:

1. During two webinars before the standard-setting meeting,
2. During the standard-setting meeting after the framework presentation but before the item review task,
3. During the standard-setting meeting before setting the round 1 bookmark, and
4. During the standard-setting meeting before setting the round 2 bookmark.

The content facilitator will be responsible for editing the descriptions drafted by the panelists at each stage and developing a coherent borderline performance description. The content facilitator will edit and modify draft descriptions at each stage to assure coherence and alignment to the framework, appropriate emphasis for each area of the framework, and appropriate calibration. The content facilitator will be responsible for communicating concerns to panel members (and the process coordinator) during working sessions should any concerns arise during the drafting/editing sessions. The process facilitator and the content facilitator will work together to develop the agenda for the webinars and to facilitate discussions during the standard-setting workshops.

In order for the replicate panels to make comparable judgments in the standard-setting workshops, they must reach a common agreement on the minimal level of performance required in the subject area to represent preparedness for eligibility for the job training or for placement in a credit-bearing college level course. The two webinars before the standard-setting meeting will be a joint meeting of the whole panel for each subject area. Activities to finalize the borderline student descriptions conducted at stages 2, 3, and 4 above, described in the section on the

standard-setting meeting, will also take place in joint panel sessions for each subject area during the workshops.

Prior to the webinars, panelists will be sent material to prepare for the discussion, including:

- Purpose of the webinars and meeting agenda for the webinars
- The NAEP Basic achievement level description for the subject to serve as a model for developing the descriptions
- Additional examples as appropriate (e.g., ACT College Readiness Standards and SAT Standards for Success)
- 12th grade student preparedness definitions adopted by the Governing Board (Appendix B)

Note that the borderline descriptions developed in the pilot study may be used as the starting point or presented during the process for developing borderline performance descriptions by the operational study panelists. The use of pilot study descriptions will be evaluated and a decision will be made by consultation among the project director, content facilitators, and the Governing Board's COR. This may reduce the amount of time needed for review and modification of the performance descriptions in the operational studies requiring an adjustment in the agenda.

BRIEFING MATERIALS FOR WORKSHOPS

Briefing materials will be provided to panelists in stages. Before the webinars and standard-setting meeting, all panelists will be sent materials that contain important background information on NAEP. Electronic communication is encouraged to the extent feasible and practicable. Briefing materials and information for the workshop will include, but is not limited to, the following:

- Cover letters with instructions for preparing for the webinars and the standard-setting study
- 2009 Mathematics or Reading Framework
- Definition of 12th Grade Student Preparedness
- Briefing Booklet describing the standard-setting process
- Meeting agenda
- Governing Board and NAEP brochures
- Confidentiality agreement
- Request for taxpayer I.D. number and certification (W-9)
- Meeting site information and transportation instructions

The timing of communications, as well as their volume and frequency, is very important to provide enough information to avoid concerns on the part of panelists but not to become burdensome. Most people seem to prefer frequent distributions of smaller quantities of materials. Therefore, the materials will be distributed in three separate communications. The first communication will be sent soon after each panelist commits to participating. Once the panels are recruited, panelists are formally invited to participate in a standard-setting workshop, including the pre-meeting webinars. A letter is sent by the project director to thank the panelists for agreeing to participate, to remind them of the dates and times of the webinars and the meeting, to inform them of how to make airline reservations for the meeting, and to provide some additional information about what they can expect.

The first packet will focus on the webinars. It will include the purpose of the webinars, the agenda and names of participants, the Basic NAEP achievement level descriptors for the

subject to serve as a model for developing the borderline preparedness description, additional examples of descriptors from other sources, and a copy of the Governing Board's definition of NAEP 12th grade student preparedness. Participants will be provided with call-in instructions and how to reach a help desk if there are problems in joining the webinar. A cover letter will explain why the webinars are important for preparing for the standard-setting meeting.

The second communication will include information about NAEP and a summary description of the standard-setting process, along with a preliminary agenda. In addition, a copy of the reading or mathematics framework will be included. A letter will explain that the emphasis of this meeting is to define standards for 12th grade preparedness for college course placement or workforce training programs, and that the framework and the definition of 12th grade student preparedness (Appendix B) will be used to help define borderline performance to represent preparedness for the postsecondary activity. Panelists will also be reminded to make airline reservations, if that has not already been done. This information will be timed to arrive about one month prior to the standard setting workshop.

The final communication will be timed to arrive about ten days prior to the beginning of the standard-setting meeting. The letter will provide detailed information regarding logistics, travel, the exact meeting site, the city, transportation to and from the airport, check-in procedures, and so forth. In addition, the packet will include a confidentiality agreement form and a W-9 form. An important part of this final communication is the "Briefing Booklet."

The briefing booklet includes a complete description of each step in the standard-setting process, in sequence, and provides the purpose for each activity as well as the methods for achieving the purpose. Combined, this set of common reference materials and introductory information related to the standard-setting process should form a sound basis for achieving a successful standard-setting study for either higher education or an occupational training program.

STANDARD-SETTING WORKSHOP

Orientation

The standard-setting workshop starts with an orientation session in which panelists will be given an overview of NAEP and the Governing Board plus a general introduction to the standard-setting method that will be implemented. This introduction will explain how the workshop fits into the overall NAEP 12th grade preparedness research program and describe the fundamental purpose and rationale for the procedures to be implemented for the workshop. Topics will include how panelists were selected, the meaning of criterion-referenced standard setting, and general themes that account for the various presentations, tasks, and exercises in the meeting. The themes will be:

- understanding the context of judgmental standard setting for higher education or occupations,
- understanding the NAEP assessment and student performance,
- understanding the definitions of student preparedness and borderline performance,
- understanding the tasks performed in recommending cut scores,
- performing the tasks to recommend cut scores, and
- evaluating the process.

The mathematics and reading panelists will meet jointly for the orientation, and NAEP test-taking. Following those joint sessions, mathematics and reading standard setting will be held separately. Within a subject, the two replication panels will meet in separate rooms for the remainder of the meeting and come together only for reaching common agreement on the borderline performance description of preparedness.

General Orientation

In a brief welcome and introduction session, the project director will introduce the staff, the process and content facilitators, and the observers to the panelists. In addition, the process for selecting panelists will be described.

Following the welcome and introductions, the Contracting Officer's Representative (COR) for the Governing Board will provide panelists with background information on NAEP and the Governing Board. This session will briefly cover the history, organizational structure, procedures, and key policies of the NAEP, as well as the purpose of setting standards for 12th grade preparedness. Information about the development of and the achievement level percentages for the 2009 NAEP grade 12 mathematics and reading assessments will also be presented.

Orientation to the Modified bookmark Method and Materials

Panelists will next be given an orientation to the modified bookmark method to be implemented for setting cut scores to represent preparedness for each subject. The purpose of this orientation is to give panelists a general overview of the specific process and to prepare panelists for the item review tasks in round 1. More complete details regarding the training provided to panelists may be found at <http://www.nagb.org/publications/2006-q12th-econ-process-report.pdf>.

First, the process facilitator provides the panelists with a basic overview of the procedure. Panelists are given a review of the framework and the preliminary borderline performance level descriptions developed via the webinars. Panelists are informed that the process includes three opportunities to place the bookmark to represent the minimal level of performance representing preparedness for the specific postsecondary activity of the workshop. Each round is designed to provide information and feedback to inform the judgments required for setting cut scores on NAEP to represent preparedness. The primary steps in the process are listed. In the first round, panelists will review the assessment framework, take a form of the test, review the items in their item rating pool, and then discuss and possibly revise draft descriptions of borderline preparedness performance for NAEP in the subject. During round 1, panelists will become very familiar with the items in the assessment by identifying the knowledge and skills that a student needs to answer each item correctly. In the second round, panelists will review actual examples of students' scored test booklets, illustrating different levels of student performance on the NAEP scale. During this review, panelists will evaluate how well performance in each test booklet illustrates the minimal level of preparedness required for the postsecondary activity of the study. Panelists will then determine if the cut score should be raised or lowered in order to better represent the minimal level of preparedness required. In the final round, panelists will be presented with impact data. Examples of data that might be used as impact data include the percent of students above the cut score on the 2009 mathematics or reading assessment and data that might be available from other studies of postsecondary preparedness. The use of any non-NAEP data as feedback must first be approved by the Governing Board, however. This information is provided to give panelists a basis for evaluating the reasonableness of the cut

scores prior to making their last judgments regarding the cut score to represent the minimal performance required to be prepared for the postsecondary activity of the study.

Next, to prepare panelists for the item review task in round 1, information on the number and types of items on each assessment is provided, followed by training on key materials and concepts used during item review—the Primary Item Map, the response probability (RP) criterion, and the Ordered Item Book (OIB). Panelists will first be introduced to the item map. Figure 2 shows a slide of a simplified item map. This slide was used in the 2009 NAEP Science ALS project to explain the general principle of an item map as spatially representative of achievement measured on the NAEP from low to high. The assessment items are located on or mapped to the scale score at which the probability of correct response is .67.

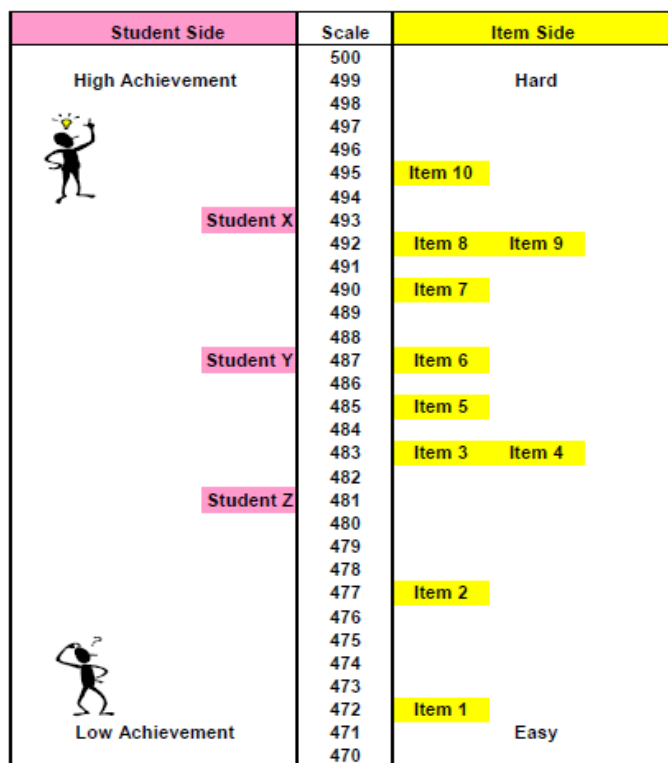


Figure 2: A simplified item map as spatially representative of achievement from low to high

Figure 3 shows a slide used to explain the role of the response probability criterion (RP criterion) in determining the location of items on the map. This slide explains the location of Item 5 in Figure 2 as a function of the probability of a correct answer on that item at a given score point on the assessment. Items are mapped to the assessment scale values based on an RP criterion of 0.67. In other words, an item is mapped to the scale value at which a student has a 0.67 (or 67%) chance of answering the item correctly. This RP criterion will be used to define *mastery* and panelists will be instructed to consider a 2-in-3 chance as meaning mastery of the relevant content reflected in the item. Introducing this concept early is important in helping panelists understand this criterion and take it into account in their bookmark placements.

Response Probability Function of Item 5

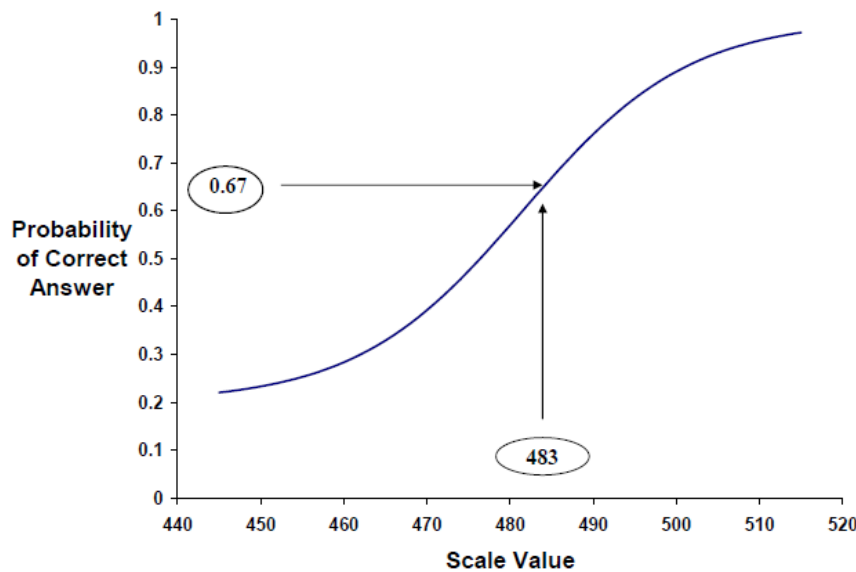


Figure 3: The relationship of the RP criterion to an item's scale value

Panelists will then be shown a Primary Item Map such as the map illustrated in Figure 4 (example from the 2009 NAEP Science ALS project). A different set of scale values will be used for constructing the Primary Item Map for each separate replicate panel. Separating the items into content-related columns (in this case, Physical, Life, and Earth & Space) provides the panelists with a layer of organization when they look at the map. This allows them to see which items measure a related set of skills (skills within a content area) and to think about what makes one item more difficult than another within a content area. The item map also illustrates the distribution of all of the assessment items on the achievement scale, mapped from easiest to hardest. Panelists are shown how this map will allow them to compare differences in difficulty between items by identifying the distance between those items on the map. A slide like that in Figure 4 is used to show how the different types of items (i.e., multiple-choice, short constructed-response, and extended constructed-response) are represented on the Primary Item Map.

Content Areas and Item Handles (Grade 4 Item Map)

Primary Item Map Grade 4

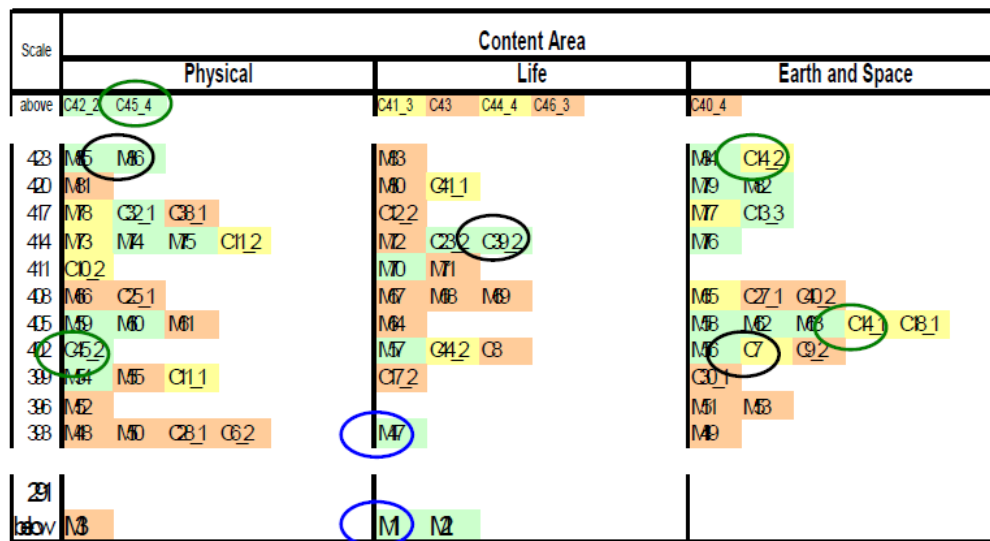


Figure 4: Illustration of how items are displayed on an item map

Each item is represented on the map by a handle—a unique identifier—consisting of a character followed by a number (e.g., M1, C1, C39_2). The first digit of the handle represents item type (C = constructed response and M = multiple choice). The number following the character represents where that item falls in order of difficulty within type. For example, M1 is the easiest multiple choice (MC) item and C7 is the 7th most difficult constructed response (CR) item on the grade 4 science assessment. The difficulty rank of each item is based on the difficulty of receiving full credit (the last or highest score level) on the item.

The scoring of extended CR items allows for partial credit. For example, on a two-point extended CR item, a student whose response is partially correct will get one point and a student whose response is fully correct will get two points, or full credit. Extended CR items occur in multiple places on the item map, one place for each possible score level. Handles for extended CR items include an underscore “_” followed by the score level. Short CR (i.e., 1-point or dichotomous) items only have 1 score level, so their handle does not include a dash and number.

The score locations of C14, a two-point CR item, are circled on the map in Figure 4. The scale value for the first score point, C14_1, is in the map score interval with midpoint 405; and the scale value of the second score point, C14_2, is in the score interval with midpoint 423. Short CR items can be distinguished from extended CR items by their handles. C7 is an example of a dichotomously-scored item.

As stated earlier, the item pool will be divided into two groups—one for each replicate panel--group A and group B. The color of an item handle on the map indicates whether the item is in the group A pool only (tan), the group B pool only (green), or in both item pools (yellow). Item C14 was in both item pools. Items in both pools are *common* items.

Panelists are then oriented to the Ordered Item Book (OIB), which accompanies the Primary Item Map. The OIB contains all of the items with which the panelists will be working in order of their difficulty, beginning with the easiest. Figure 5 illustrates this concept.

The Ordered Item Book contains test items ordered by their *scale values*, from easiest to hardest, based on student performance data.

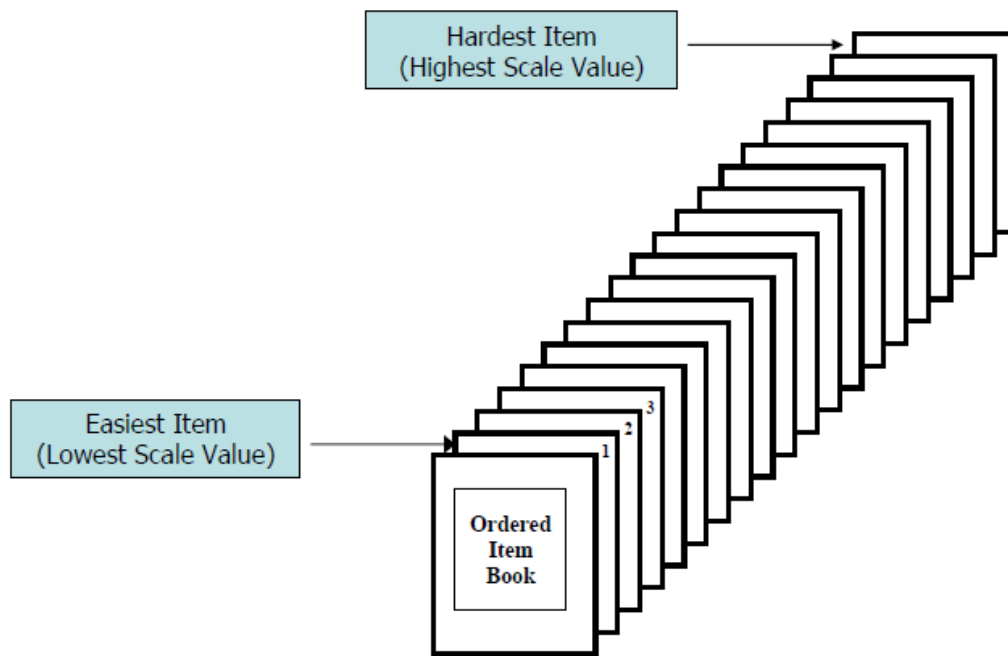


Figure 5: Illustration of how items are ordered by difficulty in the Ordered Item Book (OIB)

The slide in Figure 6 shows the location of the two score points of item C10 in the group A and group B OIBs and indicates the information contained in the OIB for each score point. Score points of extended CR items are treated as separate items in the OIB, just as they are on the item map. In the group A OIB, the first score point of item C10 was located on page 38 and the second score point was located on page 79. There are at least two pages for each score point of a CR item in the OIB—one showing the item and one showing the scoring rubric—but the page numbers in the OIB increase only when the item or score level change.

Item C10 (Grade 4 Example)

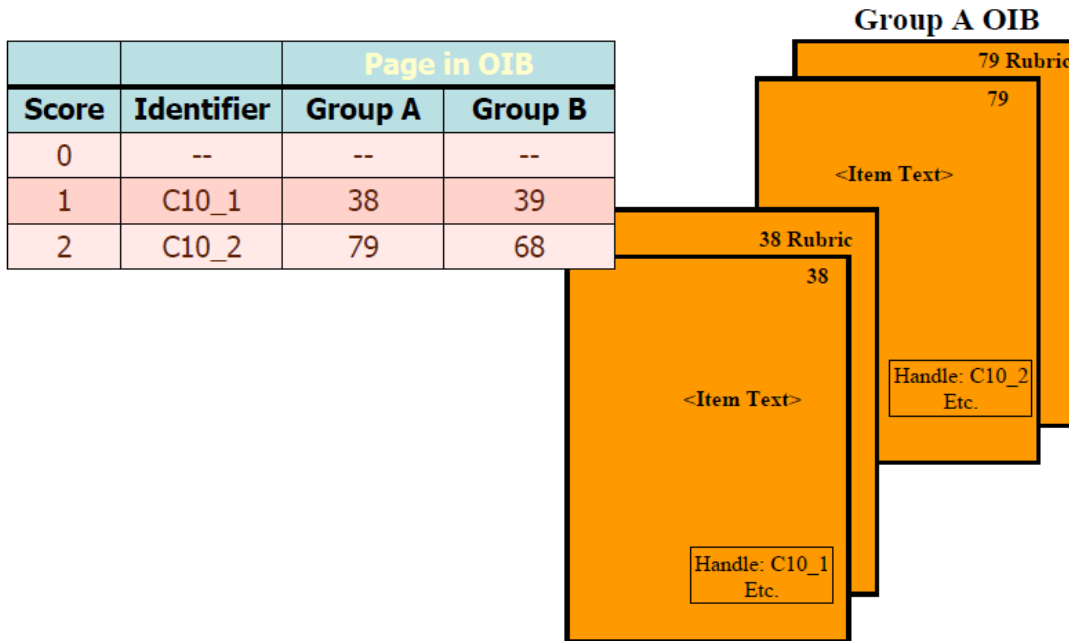


Figure 6: Illustration showing item location and information location for Item C10 in the OIB

On the OIB page that contains the item's text, there is a framed box, as shown in Figure 7. The information box will be brought to panelists' attention and the information explained. The box contains the following information for the item or score point:

- handle,
- scale value (the scale value at which a student has a 0.67 probability of earning the score point or correctly answering the item),
- map value (the midpoint of the interval containing the item on the item map),
- content area categorization in the assessment,
- other major categorizations in the assessment,
- answer key,
- identification code, and
- block and sequence number.

Item Example

OIB Page: 50

1. Cathy's garden produced ten pounds of tomatoes, five pounds of onions, and twenty pounds of cucumbers. If Cathy sold the tomatoes for 16 cents per pound, the onions for 65 cents per pound, and the cucumbers for 25 cents per pound. How much did her produce sell for?
- A. 12.30
 - B. 16.30
 - C. 9.85
 - D. 8.75

ITEM HANDLE:	M46
SCALE (MAP) VALUE:	391 (390)
CONTENT:	Mathematics
PRACTICE:	Identifying Principles

ANSWER KEY:	C
ACCNUM:	VC341275
BLOCK SEQUENCE:	3, 11

Figure 7: Illustration of the information on an OIB page

A slide with information included in Figure 8 will then be presented to briefly describe the bookmark placement process for setting the round 1 cut score. The facilitator will explain that the setting of the cut score will be criterion-referenced, not norm-referenced, based on the description of the borderline performance. It will be explained that the role of panelists is to judge the level of performance that *just* meets the preparedness requirement for an entry-level college credit course or an occupation's training program.

Round 1 Bookmark Placements

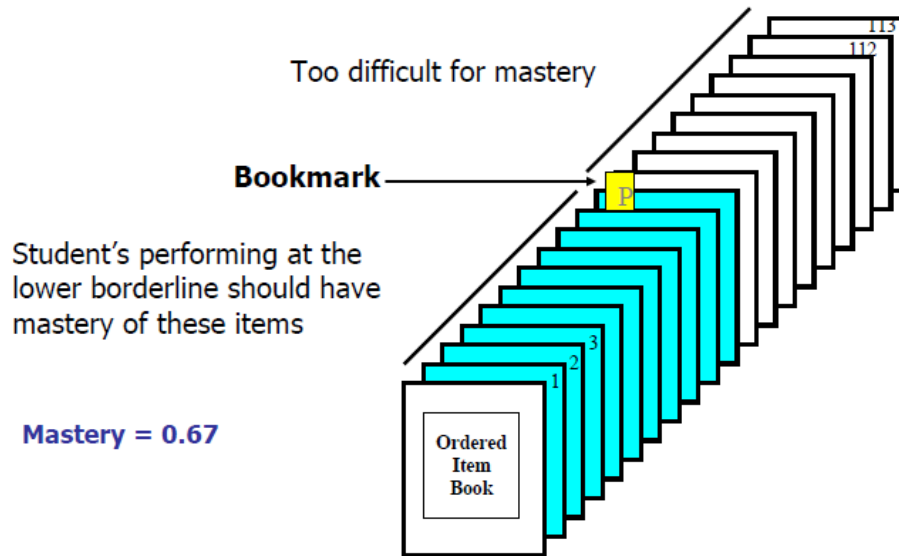


Figure 8: Illustration of round 1 bookmark placement

Round 1: Understanding the Assessment and Defining the Borderline Student

Overview of Round 1

The panel members will introduce themselves to their colleagues in each replicate panel group for a subject. And, when the two replicate groups meet jointly, they will introduce themselves to the members of the entire group. Introductions must be very brief, however.

The first round of the standard-setting method includes taking an NAEP test, a discussion of the assessment framework, review and revision of the borderline student description, a review of the items with identification of what a student needs to know and be able to do to get an MC item correct or to score points on a CR item, a second review of the borderline student description, and the placing of the bookmark for the borderline student.

Taking NAEP

Round 1 will start with panelists' taking a form of the NAEP which includes the common block that each replicate panel for a subject reviews. The panelists will take the test under test-administration constraints similar to standard conditions for the NAEP. The purpose of this exercise is for panelists to gain some insight into what students experience taking the test. After completing the test, panelists learn that their tests will not be scored or used in any other way during the meeting. After completing the test, panelists will be given training in how to use the scoring rubrics for constructed-response items. They will be provided with scoring guides and given time to score their own responses. This exercise provides an opportunity for panelists to become familiar with assessment items and scoring rubrics for items.

The Assessment Framework

To describe the knowledge and skills of a borderline student and to set a cut score on an assessment, one must have a good understanding of the assessment and of the knowledge and skills the assessment requires students to demonstrate in order to earn successively higher scores on the test.

Panelists will have been instructed to read sections of the assessment framework prior to the meeting, and they will have worked with the framework to draft the descriptions of borderline performance in the webinars. To reinforce this learning, the framework presentation provides a clear, comprehensive account of the content and organization of the framework for grade 12. This review is intended to orient panelists to the knowledge and skills that the framework covers and the specific terminology used. This discussion will help relate the draft borderline descriptions of preparedness to the assessment framework and stimulate further consideration of the appropriateness of the draft descriptions.

Review and Revision of the Borderline Student Description

The content facilitator will provide the panelists for a subject area the draft description of the borderline student that was prepared during the webinars. Based on their new experience of taking and scoring NAEP and from the detailed discussion of the assessment framework, panelists will bring additional insights to refine the descriptions. Panelists will discuss and decide upon additional revisions as a single subject-panel group.

Item Review

The second step in obtaining a good understanding of the assessment and the knowledge and skills the assessment requires students to demonstrate in order to earn successively higher scores on the test is to review test questions. Panelists will divide into the replicate panels and will spend a significant amount of time (see the draft agenda in Appendix A) identifying the content knowledge and skills that students need to know and be able to use in order to earn full credit on successively more difficult items on the test. There are four stages to this activity.

- **Stage 1** – Panel review of selected common CR items. This is a group activity conducted across replicate panels for a subject group discussion, led by both the content and process facilitators, in which panelists are trained in the process of identifying mathematics or reading content and knowledge required by CR items. The content and process facilitators will model the item review task for a sample of about four items in the common block that illustrate the various types of scoring rubrics associated with the CR items. They will begin with an easy item and progress to more difficult items. For each CR item, they will identify and make notes on what students need to know and be able to do to get full credit on the item; then they will identify and make notes on the knowledge and skills needed to earn successively lower scores on the item.
- **Stage 2** – Small group review of the remaining CR items by panelists in separate replicate panel groups. This item review task will be conducted in small groups for each replicate—panel A and panel B. Each small group will implement the process modeled in stage 1 to review the remaining CR items in their item-rating pool. Following group discussion of the content and knowledge required at each score point of an item, panelists will make notes as to the knowledge and skills they judged necessary to earn successively lower scores on the item. Panelists will take turns “leading” this activity in their group. Content and process facilitators will circulate to answer questions and provide guidance as needed. Since borderline performance is unlikely to be at the

extremes of the score range, it is proposed that panelists' review be focused on the items in the range from the median of the scores below the Basic cut score to the Advanced cut score for the 2009 grade 12 NAEP in the subject. Prior to the meeting, the contractor will submit a plan for approval by the COR to spiral the items in this range across the panelists so that each of the items in the rating pool for a replicate panel group is reviewed by at least two panelists. This spiraling plan will determine the time frame of the item review session. The panelists will be instructed to read through the materials for all CR items, but to only write notes for the items included on their CR Item Review List.

- Stage 3 – Independent review of OIB. This is an independent review task in which panelists identify the knowledge and skills required by multiple-choice items in their pool in the context of their OIB. They will consider all items in their pool sequentially, beginning with the first, or easiest, item. An important part of this task is to think about the additional knowledge and skills that an item might require that was not required by earlier, easier items representing similar content. During the independent review of the OIB, panelists will make notes on what students need to know and be able to do to answer each MC item correctly and they transfer their notes on CR score points to the OIB as they encounter each score point in the OIB. This review enables panelists to become familiar with the progression of difficulty from one item to the next within their item-rating pool. As for the CR items, the review will be focused on items from the median of Below Basic to the Advanced cut score, although items from across the entire score range will be included in the review. Prior to the meeting a plan for item review will be developed. Panelists will be given a list of MC items for which to write notes. The panelists will be instructed to read all MC items, but to write notes only for the items included on their MC Item Review List.
- Stage 4 – Table-group discussion of OIB. Next is a table-group discussion of the knowledge and skills associated with each item/score point in the context of the OIB. Again, items are considered sequentially, beginning with the easiest. As before, very easy and very difficult items will be sampled and not all the items in the pool will be discussed. Panelists will share their ideas about the knowledge and skills needed to get each item correct and add to their notes.

Materials for stages 1 and 2 of the item review include the Constructed Response Ordered Item Book (CROIB), a CR Item Review list, and a Notes template. The CROIB contains all the CR items in a group's item pool. Items are listed in order of difficulty by the highest score point.

The slide in Figure 9 illustrates the contents of the CROIB. Unlike the OIB, all the information about a constructed-response item is contained together, on consecutive pages within the CROIB. Items are separated by tabbed pages, with the tab showing the item handle (minus the score points). Item information includes the scoring rubric and examples of student responses at each score level, including zero. The first page shows the item, the information box, and the page number(s) where the item's score point(s) can be found in the OIB.

Constructed Response Ordered Item Book (CROIB)

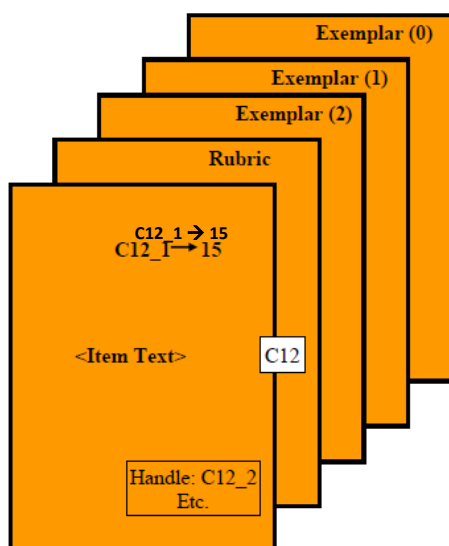


Figure 9: Illustration showing information location for Item C12 in the CROIB

Because the panelists will need to record the knowledge and skills identified by going through the CROIB and then adding their notes from the CROIB to the OIB, panelists will find it helpful to use large yellow Post-it Notes to record their notes on each CR item. The notes are for the individual panelist and need only to be informative to the panelist. A separate description is needed for each score point for each CR item. When panelists are finished with an item, they place their notes in the Notes template. The Notes template is a stapled set of 11x17 pages with outlines for accommodating ten Post-its per page. Within each outline is an item handle and OIB page number identifying the Post-it that is to be placed there. The contractor may propose a computer-based method for recording the item descriptions and associating those with the items in the OIB. This note-taking step is labor and time intensive, and a computer-based method might speed the process of both taking notes and associating the notes with the items in the OIB.

During stage 3, panelists make notes on what students need to know and be able to do to get the MC items correct. Because of time constraints, each panelist will review and make notes on about 60% of the MC items in the OIB, primarily those that map from the median of Below Basic to the Advanced cut score for the 2009 grade 12 NAEP in the subject. To ensure that all MC items likely to be in the range of the cut score are reviewed, each panelist will be given a list of specific MC items to review, and all MC items in the item pool will be available for review if time permits.

As panelists progress through their OIB, they will transfer their notes on CR score points from the notes template to the corresponding OIB page as they encounter each score point. As noted earlier, the OIB contains all items, including the constructed response items. Figure 10 shows how score levels of extended CR items are treated as separate items in the OIB. The use of the notes template will allow panelists to place their notes on the scored item steps on the correct

OIB page numbers with just one pass through the OIB. This will allow panelists to see their constructed response item notes in the context of all of the items in the OIB. When panelists see score points of extended CR items relative to the difficulty of all other items in their pool, they will be able to add to their notes observations about what content knowledge and skills the score point may require that previous, easier items and score points did not require. Panelists may record further notes directly on the pages of the OIB.

Panelists will also check MC and CR items off on their Primary Item Map as they progress through the OIB. The item check-off process helps panelists see “how much” more difficult one item is than another and which items are related in terms of the general knowledge and skills that distinguish different content areas.

Constructed Response Items Are Treated As Separate Items and Appear at Different Places in the OIB

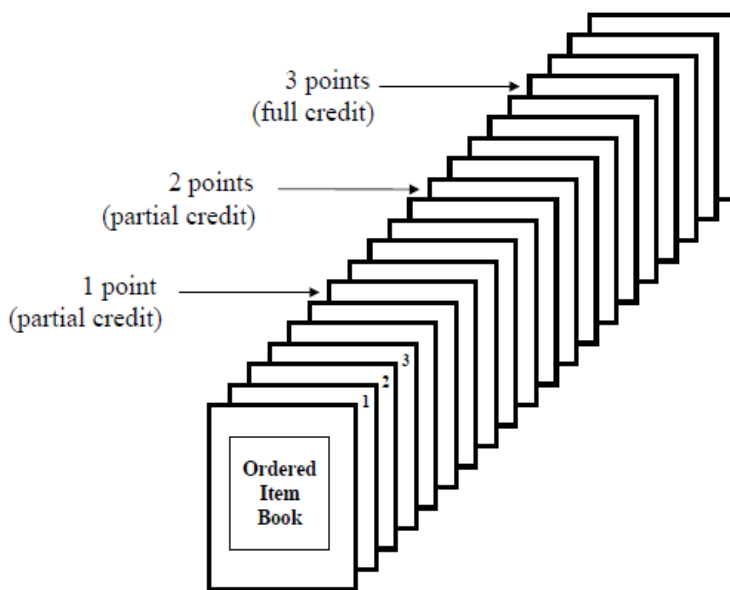


Figure 10: Illustration showing the location of extended CR item score levels in the OIB

In the table-group discussion (stage 4), panelists share their ideas about what students need to know and be able to do and add the ideas of other panelists they agree with to their notes. Panelists take turns leading the table discussion. The process is monitored by facilitators to assure that all panelists contribute to the discussion process.

When the item review is complete, panelists will have a detailed, structured understanding of the assessment and student achievement. Structure is provided by the difficulty-order of content knowledge and skills required by test items as shown in the OIB and on the Primary Item Map. This structure prepares panelists to understand the continuum of increasing knowledge and skills represented by the increasing scores on the achievement scale.

Borderline Student Description

After reviewing the items in each replicate panel, the subject group of panelists will come together again to discuss the description of borderline performance. After looking at the items, are there other ways of describing the knowledge and skills that should be incorporated into the description of borderline performance? Are there areas in the description that are unclear? The content facilitator will lead the discussion and note appropriate revisions in the performance level description prior to the first bookmark placement. All panelists in the subject group must reach agreement on the minimal level of performance required to represent preparedness for placement in a college credit course or a job training program. Having panelists discuss this performance with one another has been an effective means of helping them to solidify their understanding of the performance requirements.

Placing the Bookmark

Once the description of the borderline student has been reviewed and revised as needed, the bookmark placement task (or standard setting) will begin. The structure provided by the OIB and Primary Item Map prepare panelists to apply the borderline performance description when placing their bookmark.

The bookmark placement task will initially be described to panelists as a process of going through the OIB, beginning with the easiest item, until they come to an item that they judge to be too difficult for “mastery” of the minimally prepared student. Mastery will be defined as having at least a 0.67 (2/3) probability of answering the item correctly. This is referred to as the response probability (RP) criterion. The bookmark is placed on the item immediately preceding the item judged to be too difficult. Figure 11 illustrates what is meant by “mastery” of the items at and below the cut score—mastery means that a student performing at the specific scale score has a 0.67 probability of answering the item at that score correctly, a higher probability of answering items below the score correctly, and a lower probability of answering items above the score correctly.

It means this!

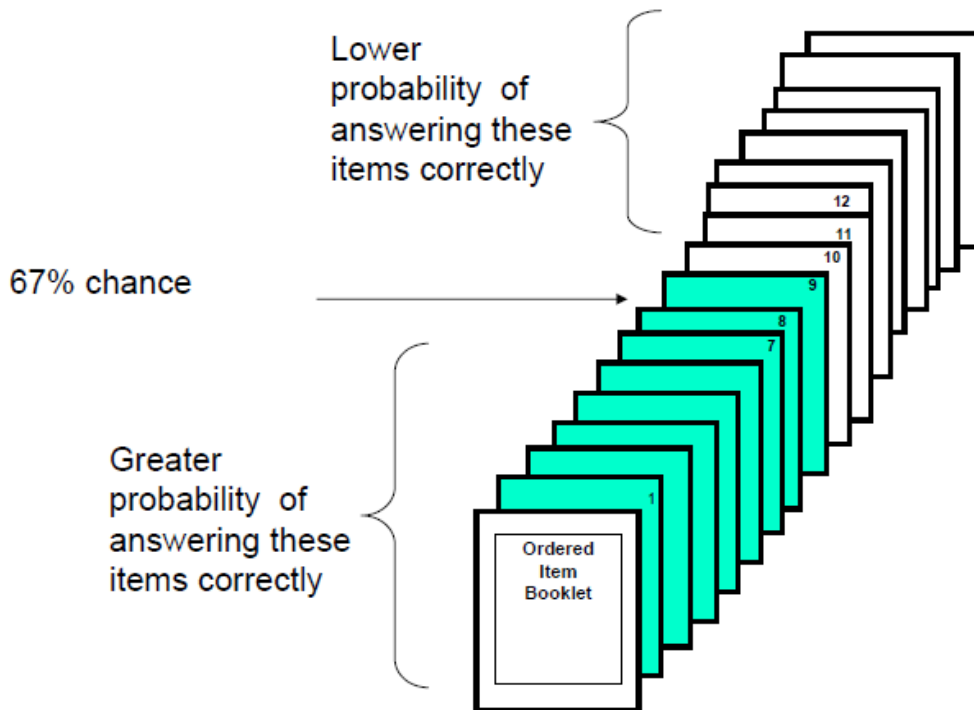


Figure 11: Illustration of the relationship between bookmark placement and the “mastery” of items at and below the cut score using 0.67 as the RP criterion

Once panelists understand this idea, the process facilitator will explain to panelists that it is possible for them to be unsure of where to place their bookmarks because (a) they may not feel there is a noticeable or meaningful difference between adjacent items in terms of difficulty, and (b) they may feel that a few items in the OIB are out of order with their own expectations of relative difficulty.

The initial description of the process is then supplemented with the instruction to go beyond the first item they judged to be too difficult to see if any of the items that follow should have been mastered by students who just meet the description of borderline preparedness performance. This instruction will be represented to panelists visually by showing a *range of uncertainty* in a slide depiction of the OIB. All items below this range are judged to have been mastered items. All items above this range are judged *not* to have been mastered items. Figure 12 shows an illustration of this concept used for panelists in previous NAEP standard-setting projects. Panelists are told that any item in the range of uncertainty would be an acceptable choice for placing the bookmark. Panelists should be reminded to refer to the Primary Item Map to supplement their evaluation of the items and the placement of the bookmark to determine the relative difficulty of the items being considered for the bookmark placement. If panelists cannot distinguish a clear choice for placing the bookmark, they may choose the middle of the range of uncertainty.

Identify and focus on items whose mastery/nonmastery status you are unsure of with regard to borderline Proficient

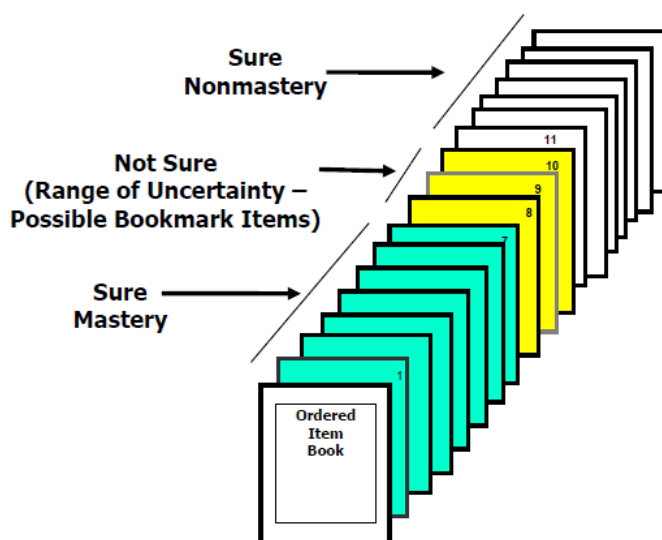


Figure 12: Illustration of the range of uncertainty in a bookmark placement

The process facilitator will briefly review the bookmark placement task and ask the panelists to summarize the assignment in their words. Any questions will be answered. Panelists will be instructed to place their bookmark independently, with no discussion in their group.

Panelists record the page number of their bookmark on a Cut Score Recommendation form. They will circle the handle of their bookmarked item on their Primary Item Map. Staff will subsequently record the scale value corresponding to the bookmarked page beneath the bookmarked page number on the panelist's form and compute the median cut score for the panel.

Round 2: Understanding Student Performance

Overview of Round 2

The second round of the modified bookmark method begins with a presentation of the cut score results from round 1. Panelists then receive holistic feedback in the form of actual student test booklets to help with their understanding of what students *can do* with respect to the NAEP in the subject. Panelists are led to examine performance right at the cut score computed for the round 1 bookmark, as well as above and below that level. Following review and discussion of the student booklets, panelists will discuss and revise the description of borderline for the last time. Panelists then select a scale value to represent the level of performance to just meet the requirements described for borderline preparedness.

Feedback from Round 1

At the beginning of round 2, results from round 1, including the median cut score and the distribution of cut scores, are described. The scale used for reporting results will be different for each separate replicate panel. Figure 13 shows an example of the cut score distribution chart

that will be provided as feedback from round 1. This chart is used to illustrate the location of all panelists' round 1 cut scores. It helps individual panelists to evaluate the location of their cut score relative to that of others. Panelists who have placed their cut score far from those of most other panelist should attempt to ascertain how their conceptualization and understanding of borderline performance required for preparedness differs from that of others in the group.

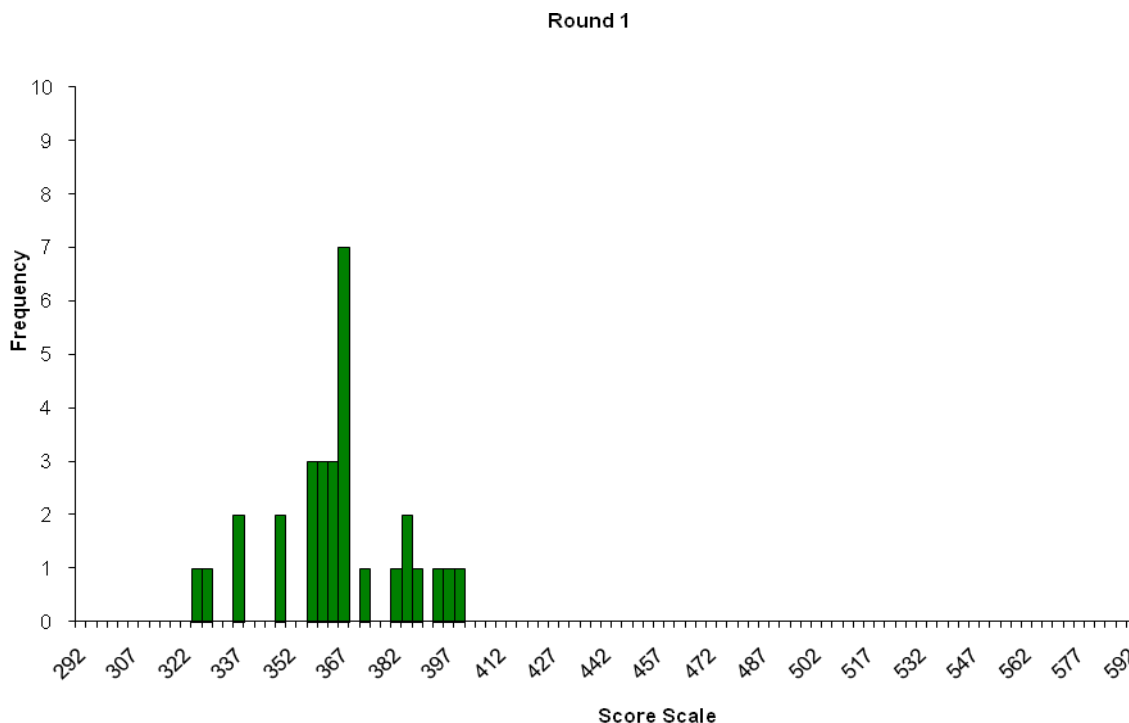


Figure 13: Illustration of cut score distribution chart showing the distribution of panelists' cut scores

In addition to providing the numerical value of the cut scores, feedback is shown on the item maps. Panelists are given a new version of their Primary Item Map with the panel median cut score on the map as shown in Figure 14. Panelists are instructed to circle their round 1 bookmarked item on the item map so they can compare the panel cut score and bookmarked items to their own cut score and bookmarked items. They are again reminded to evaluate the relative location of the items on the Primary Item Maps.

Panel Cut Score on the Primary Item Map

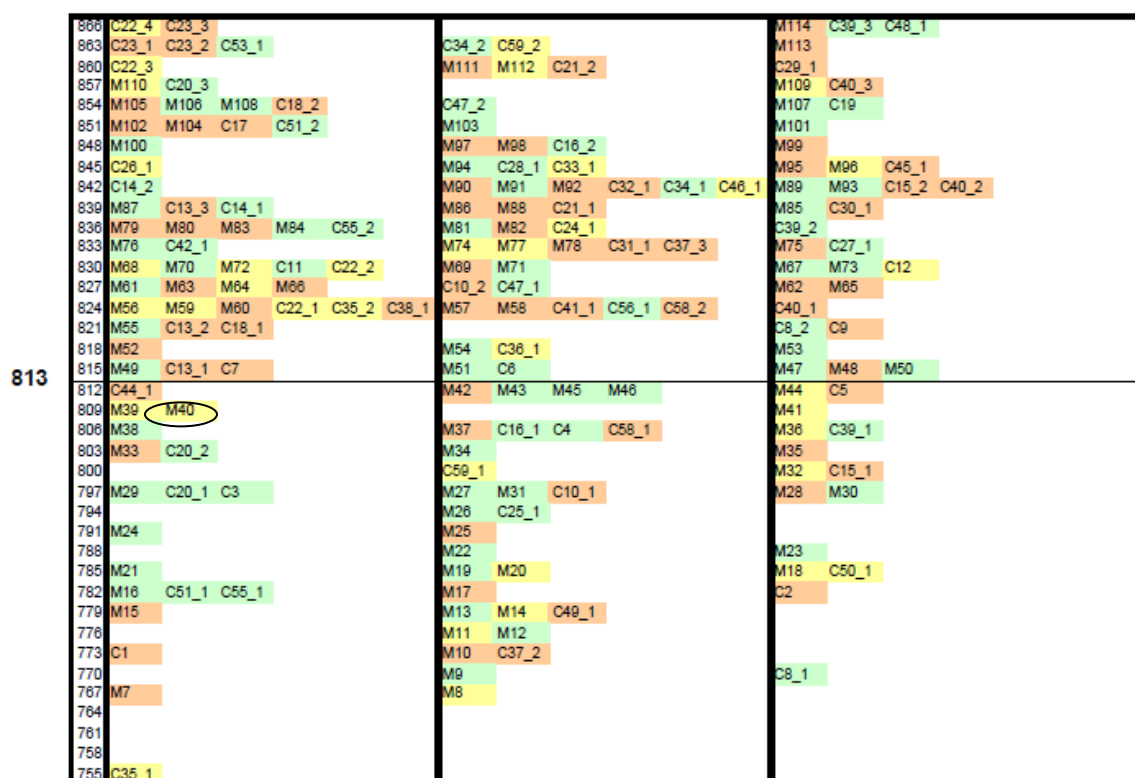


Figure 14: Illustration of Primary Item Map showing the round 1 panel cut score (horizontal line) and the location of a panelist's bookmarked item (circled)

Panelists are instructed to flag (using a small Post-It, for example) the panel cut score in their OIBs. To focus their attention on the intended, criterion-referenced meaning of the round 1 cut score, panelists are instructed to identify the items that fall between their cut score and the panel's median cut score and to determine what these items represent in terms of differences in performance between the two cut scores, as illustrated in Figure 15. Panelists are cautioned to recall the location of their cut score in relation to cut score for the entire panel because examples of student performance will be provided to represent the cut score for the entire panel, computed as the median of the cut scores across panelists, and not the individual panelist's cut score.

Notice location of your bookmarks relative to group cut scores

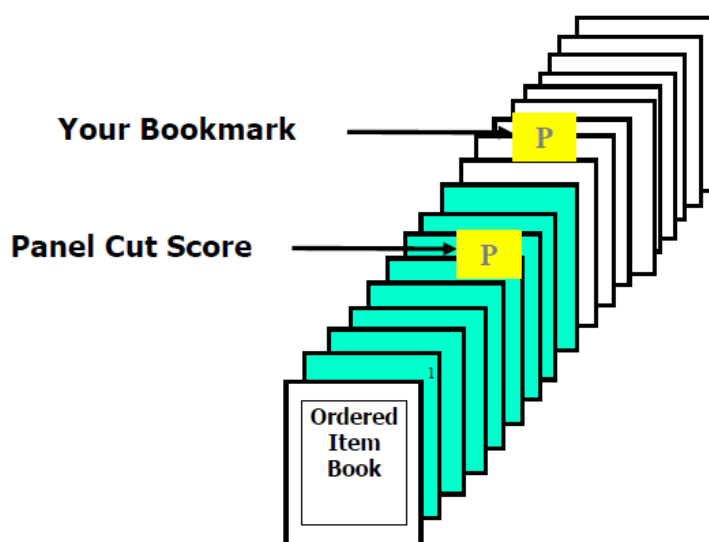


Figure 15: Illustration of the comparison of the panel cut score and a panelist's bookmarked item in the Ordered Item Book

Whole Booklet Feedback

Next, panelists are instructed in how to use the performance exhibited in student booklets scoring at the borderline for making a judgment for placement of their round 2 cut score. Panelists need to determine whether the performance represented in the booklets seems about right or is too low or too high to reflect their understanding of borderline performance. The panelist must note the location of their own cut score relative to the group cut score and relative to performance in any booklet that closely represents their understanding of borderline performance.

Six booklets on each of three forms will be provided to panelists for each NAEP subject, with each group (A and B) reviewing two forms for a total of 12 booklets per replicate panel. Each panelist will review one form that is common to both groups A and B. Booklets for review will be selected and copied prior to the meeting. From the wide range of potential booklets for each form (see section on *Document Preparation for Meeting*), booklets for panel review will be selected so they are distributed around the round 1 median cut score, with two booklets on each form scoring close to the cut score, two booklets at intervals below the cut score, and two booklets at intervals above the cut score. The booklets will be selected to span approximately plus or minus 30 points surrounding the panel cut score.

A Booklet Score Plot such as presented in Figure 16 will be prepared for each form. For each form, the expected number of points for each scale value is plotted on the Booklet Score Plot and the booklets are indicated on the plot at their scale value. These plots are used to provide a visual illustration of the location of each booklet relative to the cut score and the pseudo-NAEP score scale.

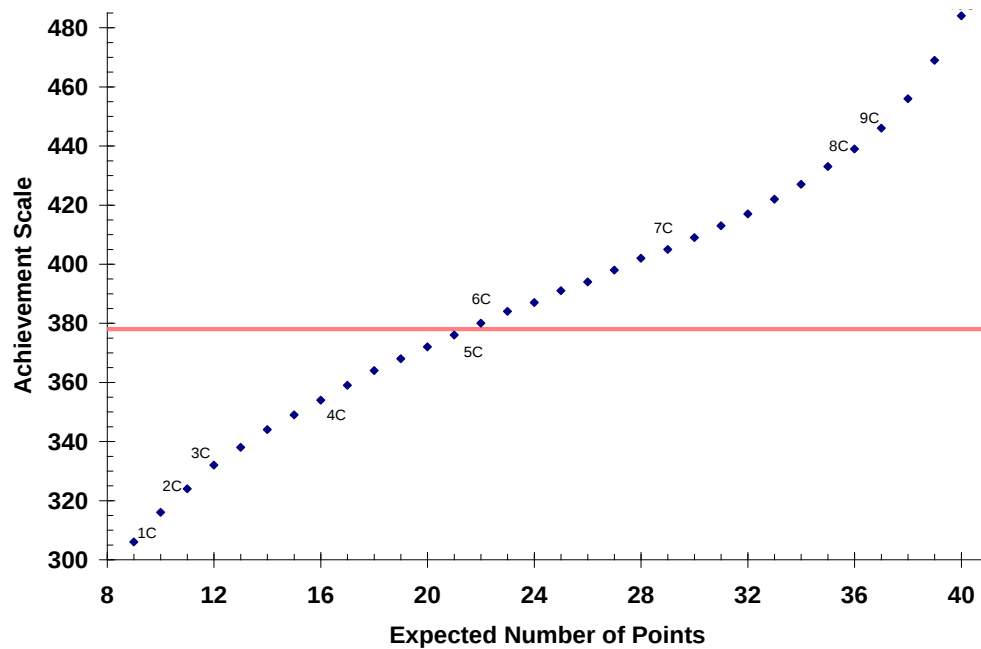


Figure 16: Illustration of Booklet Score Plot for a form, showing the round 1 cut score (horizontal line) and the score of each student booklet (1B through 6B) on the achievement scale

The Booklet Score Charts map the expected number of points correct on the common and group-specific forms to the achievement scale within a range from 10 points below the lowest panelist's cut score to 10 points above the highest panelist's cut score from round 1. The placement of the booklets on the chart is determined by their expected number of points correct. Panelists are asked to circle their cut score on the Booklet Score Chart and to take note of where their cut score falls in relation to the booklets they will be reviewing (see example in Figure 17).

Preparedness

	Scale	Common Form		Group A Only Form	
		Booklet	Expected No. of Points	Booklet	Expected No. of Points
	405		29.0		31.0
	404				
	403		28.5		30.5
	402		28.0		
High	401				30.0
	400		27.5		
	399				29.5
	398	6C	27.0	6A	29.0
	397				
	396		26.5		
	395				28.5
	394		26.0		
	393		25.5		28.0
	392				
	391		25.0		27.5
	390				
	389		24.5		27.0
	388				
	387		24.0		26.5
	386				26.0
	385		23.5		
	384		23.0		25.5
	383				
	382		22.5		25.0
	381				
	380	5C	22.0		24.5
	379				
	378		21.5	5A	24.0
	377				
	376		21.0		23.5
	375				
	374		20.5		23.0
	373				
	372		20.0		22.5
	371				
	370		19.5		22.0
	369				
	368	4C	19.0		21.5
	367				
Median →	366		18.5	3A,4A	21.0
	365				
	364	3C	18.0		20.5
	363				
	362				20.0
	361		17.5		
	360				
	359		17.0		19.5
	358				
	357		16.5		19.0
	356				
	355				18.5
	354		16.0		
	353				
	352		15.5		18.0
	351				
	350				17.5
	349	2C	15.0		
	348				
	347		14.5	2A	17.0
	346				
	345				16.5
	344		14.0		
	343				
	342				16.0
	341		13.5		
	340				
	339				15.5
	338		13.0		
	337				
	336				15.0
	335		12.5		
	334				
	333				14.5
	332	1C	12.0		
	331				
	330			1A	14.0
	329				
	328		11.5		
	327				13.5
	326				
	325				
Low	324		11.0	32	
	323				13.0
	322				
	321				
	320		10.5		12.5
	319				

Figure 17: Sample Booklet Score Chart for group A showing the median (yellow highlight), high, and low cut scores (horizontal line) and the location of a panelist's round 1 cut score (circled)

The Booklet Score Charts are specific to each replicate panel. These charts map the expected number of points on a form to the achievement scale within a range from 10 points below the low cut score to 10 points above the high cut score from round 1. The booklets are then indicated at the location of their expected number of points. A yellow highlight indicates the median cut score from round 1. Panelists are asked to circle their cut score on the Booklet Score Chart and to take note of where their cut score falls in relation to the booklets they will be reviewing (see example in Figure 17).

For each test form, the Item Score Tables (IST) provide the score a student received (0 = incorrect, 1 = correct) for every score point on each student booklet. The items and score points are ordered from easiest to hardest, bottom to top, and the student booklets are ordered from lowest to highest scoring left to right. Figure 18 illustrates the Item Score Table for Form C. Panelists can use the IST to see, at a glance, the response patterns of students across the range of the scores around the cut score. For example, in Figure 18 panelists can see that in one of the borderline cut score booklets, booklet 3C, the student received credit for about 75% of the total points and correctly answered many of the easy items and fewer of the hard items.

Before panelists begin their independent review of the student booklets, they will be led through a whole group exercise to familiarize them with the Booklet Score Charts (BSC), Item Score Table (IST), and booklet item maps. This exercise will help them begin to understand the relationship between general performance on a form of the test and expected performance on individual test items.

The panelists review the Booklet Score Charts and Item Score Tables in relation to the two student booklets at the median cut score from round 1 on the common form (booklets 3C and 4C in Figures 17 and 18). Using the Item Score Table, panelists are told to observe the response patterns of the two student booklets near the panel cut score (3C and 4C) and to note that:

- The students answered different items correctly and incorrectly, but the overall proportion of items answered correctly was nearly the same.
- Differences in correct and incorrect answers may be due to variation in student mastery across content areas.
- Students did not get all items below the cut score correct and all above incorrect, but the probability of a correct response increased the farther below the cut score an item was and decreased the farther above the cut score an item was.

Once panelists are able to understand and interpret the information provided in the Item Score Table, panelists are given the opportunity to independently review booklets 3C and 4C. They are instructed to take note of where their cut score falls in relation to the scores on these booklets, and to consider if performance represented by the booklets is too high, too low, or just right for the borderline student. A brief discussion is held following this review, in which panelists share their perceptions of the level of performance exhibited in the booklets as related to the performance described in the borderline student description. The purpose of the discussion is to help panelists begin the process of gaining a shared understanding of the meaning of borderline performance.

Scale				Cut Score					
				1C	2C	3C	4C	5C	6C
				24	26	27	28	30	32
Handle	Value	Block	Seq						
C21_3	476	2	9	0	0	1	0	0	1
C15_3	436	4	9	0	0	1	0	1	1
C13_2	435	4	4	0	0	0	1	0	0
M149	434	4	18	0	0	0	0	0	0
M148	431	2	17	1	1	1	1	0	1
C12_2	427	4	13	0	0	0	1	1	0
M143	419	4	3	1	1	0	1	0	0
C21_2	414	2	9	0	0	1	1	1	1
M134	408	4	12	0	0	1	1	1	1
M131	403	2	16	0	0	1	1	1	1
M132	403	4	17	1	1	0	1	0	1
M124	398	4	16	0	0	1	1	1	0
M118	396	4	7	0	0	0	1	0	0
M112	393	2	2	0	1	1	1	1	0
M113	393	2	15	0	1	1	1	1	1
M110	392	4	15	1	0	1	1	1	1
M106	391	4	14	1	1	1	1	1	1
M104	390	4	2	0	1	1	0	0	1
M101	389	2	10	1	0	1	0	1	1
M96	387	2	14	1	1	1	0	1	1
M93	386	2	8	0	1	0	0	0	0
C12_1	385	4	13	0	0	1	0	1	1
C15_2	383	4	9	0	0	0	0	0	1
C4_2	381	2	13	0	1	0	1	0	0
M81	378	4	1	1	1	0	1	0	1
M80	377	4	5	1	1	1	1	1	1
M67	368	2	11	0	1	1	0	1	0
M58	365	2	6	1	1	0	0	0	1
M60	365	4	8	1	0	1	1	1	1
M55	364	2	7	1	1	1	1	1	1
C21_1	360	2	9	1	1	0	1	0	0
M50	358	2	5	1	1	1	1	1	1
M42	353	2	12	1	0	1	1	1	1
M25	342	4	11	1	0	1	1	1	1
M20	339	2	18	0	0	1	1	1	1
C1_2	338	2	4	1	1	1	1	1	1
C15_1	325	4	9	0	1	1	1	1	1
C4_1	324	2	13	1	1	1	1	1	1
M7	320	2	1	1	1	1	1	1	1
M6	317	2	3	1	1	1	1	1	1
M5	313	4	6	1	1	1	1	1	1
C13_1	312	4	4	1	1	1	1	1	1
C1_1	290	2	4	1	1	1	1	1	1

Figure 18: Illustration of an Item Score Table for Form C, with the items listed from hardest to easiest, top to bottom, in the left most column and the student booklets listed from lowest to highest score, left to right, in the top row with the number correct score on the booklet provided below the booklet identifier (1C to 6C)

Following this discussion, panelists begin an independent booklet review of all 12 booklets provided to their panel. They will be asked to consider:

- How performance at the round 1 median cut score differs from performance above the cut score and below the cut score.
- How students at their round 1 cut score are performing in relation to students at the panel cut score.
- If performance at the panel cut score was higher, lower, or just right for the borderline student.

At the conclusion of the independent review, panelists will discuss with each other the above questions and share their reactions to the performance exhibited in the booklets. They will be instructed to share their ideas, but not to try to change the views of one another. The purpose of the discussion is to further inform the judgments for round 2 cut score recommendations.

Review and Revise the Borderline Student Description

Panelists will be given the opportunity to review the preliminary borderline student description and to make final changes based on their experience in using the preliminary borderline student description to set their bookmark in round 1 and their review of student booklets for round 2.

First, there will be a panel discussion in which panelists will discuss how readily they were able to relate their understanding of what students need to know and be able to do to get each item/score point correct to their understanding of the borderline performance description when they selected their bookmarks in round 1. Panelists will be asked to discuss, from their perspective of the item they chose as the bookmark item in round 1, what aspect of the borderline description suggested that this was the appropriate place for their bookmark. The purpose of this discussion is not to change anyone's opinion about the cut score; rather, the discussion is to better understand the rationale for different judgments and how different understanding of borderline performance influenced these judgments. Next, panelists will be asked to discuss if and how their understanding of borderline performance has changed following review of student booklets in round 2. Finally, panelists will be asked whether additional changes to the borderline performance descriptions are needed. This is the last opportunity for revising the performance descriptions to be used for making cut score judgments in rounds 2 and 3.

Round 2 Cut Score Recommendations

In making round 2 cut score recommendations, panelists will be instructed to work independently. Panelists will choose a scale value and record the scale value on their Cut Score Recommendation form. Panelists will be instructed to circle the scale value they choose for their round 2 cut score recommendation on their Booklet Score Chart and to move their round 1 bookmark in their OIB to the last item in their OIB with the scale value less than or equal to their recommended cut score.

Specific instructions will be provided to aid them in the selection of their round 2 cut score. They will be instructed to select a range of scale scores within which they are deliberating, their range of uncertainty. This range might encompass, for example, the panelist's own cut score at the low end and a booklet that they feel represents borderline performance at the high end. Once they have identified the range, they are to locate the high and low points of this range in their Ordered Item Book and Booklet Score Chart and to consider (a) what a student needs to know and be able to do to correctly answer items at-or-below the potential cut scores in the OIB, and (b) the performance associated with the potential cut scores in the booklets indicated on the Booklet Score Chart.

In considering booklets, panelists are also reminded of a number of technical considerations.

1. There are many different forms of the assessment and each form has approximately the same number of points.
2. The achievement scale represents a much larger range than the number of points on a form, there is not a one-to-one correspondence between the point values on the student booklets and the score scale.

3. These scale values may correspond to point values on different forms, however, so panelists should consider interpolating between raw score points on any given form when deciding to adjust their cut score.

Round 3: Impact Data

Overview of Round 3

The third round of the standard-setting method starts with presentation of the cut score results from round 2. Panelists will then receive holistic feedback in the form of impact data. These data will show the percentage of scores on the 2009 grade 12 NAEP in mathematics or reading that were at or above the cut score. The percentage at or above the cut score indicates the percentage of students who would be prepared for placement in the postsecondary activity. Research data may be available to provide additional information for comparing performance relative to the NAEP preparedness cut scores with performance relative to scores on college course placement tests or job training program admissions tests. The use of any non-NAEP data must be approved by the Governing Board in advance. Following a panel discussion of the data, panelists will independently select a final cut score to represent grade 12 preparedness on the NAEP score scale for the subject—reading or mathematics.

Feedback from Round 2

Feedback from round 2 will be presented using the same materials and formats that were used to present feedback after round 1. Feedback from round 2 will consist of the median cut score and the cut score distribution chart. Panelists will be given a new Primary Item Map, Booklet Score Chart, and Booklet Score Plots on which round 2 cut scores are marked. A table of the panel cut scores from rounds 1 and 2 will be presented to show panelists how the cut scores have changed over rounds. Panelists will then mark where their cut score falls in relation to the median panel cut score on the Primary Item Map, Booklet Score Chart, and the OIB.

Impact Data and Discussion

After round 2, panelists will also be given impact data. These data will fall into two categories:

1. The percent of students who scored at or above the round 2 preparedness cut score on the 2009 NAEP for grade 12 mathematics or reading, and
2. Research data that might be available and approved for use by the Governing Board relating NAEP grade 12 mathematics and reading scores to scores on various tests used for postsecondary placement (e.g., ACCUPLACER, ACT, COMPASS, SAT, or WorkKeys).

The information will be used both as feedback for round 2 cut scores, and as information to be used in setting cut scores in round 3. The process facilitator will briefly describe the results and how the impact data can be used by panelists to inform their judgments for round 3 cut scores.

The impact data will be discussed prior to panelists' making their round 3 cut score recommendations. Panelists will be reminded that the NAEP impact data are from the 2009 assessment, and they will be reminded of the features of NAEP that are unique relative to other assessments, e.g., individual student scores are not provided to students. But regardless of what students can do as illustrated by the impact data, what students should be able to do to meet the requirements of the borderline performance descriptions must take precedence. The discussion then is largely left open to panelists, but panelists must be prompted to have a full and open discussion of their reaction to the data. The data are provided to promote discussion

and to give a basis for judging the reasonableness of the cut scores to represent the minimal performance required to be considered prepared to enter the specific postsecondary activity.

Round 3 Cut Score Recommendations

The purpose of round 3 cut score recommendations is to allow panelists to adjust their cut score recommendation based on feedback after round 2. Panelists will be instructed to work independently, study the feedback from round 2, reflect on the discussion of the data, and determine whether their round 2 cut scores should be changed. If they decide to change their cut score, they are instructed to consult their Ordered Item Book and Primary Item Map to determine if the new cut score they are considering is consistent with the description of borderline performance. Panelists will then record their cut score on their Cut Score Recommendation form.

Post-Round 3 Activities

Feedback from Round 3

Feedback on the results from round 3 is given in the usual fashion. Panelists identify where their cut score falls in relation to the final panel cut score. Panelists will be given a new Primary Item Map with the final cut score derived from round 3 marked on it and a new Cut Score Dispersion Chart. They will be instructed to remove their bookmarks from their OIB and to discard those bookmarks. They are then told to move their panel bookmark to the final cut score. This is done to emphasize that the round 3 cut score is the final cut. The feedback will also include updated impact data based on the round 3 panel cut score. After receiving feedback on the results of round 3, panelists are asked to complete a Comparative Data Questionnaire.

At the start of the session, panelists will be told that the round 3 cut score will be reported to the Governing Board as the key outcome of the standard-setting meeting. It is very important that panelists understand the level of performance exhibited by students at the cut score, which is the purpose of the feedback, and that they evaluate the cut score based on the match between the criterion-referenced feedback and the description of borderline performance. For this final discussion of impact data, panelists from groups A and B for each subject will be convened together. It is important to inform panelists about the cut score recommendations of each group and to have their final evaluation of the cut score recommendation that they would make to the Governing Board be in light of full information regarding the cut scores set by both groups. No actual cut scores will be changed as a result, but the Governing Board will have these data to inform their decisions regarding the cut score.

Impact Data Questionnaire

The purpose of the impact data questionnaire is to provide the Governing Board with information about panelists' reactions to the final impact data. Since the availability of non-NAEP data is not yet determined, the sample questionnaire in Appendix C is brief and must be extended to take into account any new data that become available. Panelists will answer questions about their reaction to the comparative information. The questionnaire will allow panelists to recommend a cut score other than the round 3 panel cut score. This recommendation will be made after having information and discussion about the cut scores for the two panels in the subject group.

A Cut Score Proportion Chart (Figure 19) will be provided to allow panelists to see the relative impact of changing from one cut score to another if they want to raise or lower the cut score. This chart provides the percentage of students scoring at or above every fifth score value on the

NAEP mathematics or reading scale. (The interval could be increased or decreased from 5, but it must be the same for all studies.) The round 3 cut score is marked on Figure 19. Panelists will be instructed to use this information to determine their judgment of the final cut score they would recommend to the Governing Board. Note, however, that the assumption is that the Round 3 cut score will be recommended to the Governing Board unless a large number of panelists recommend a different score when given this final opportunity.

2009 NAEP Grade 12 Mathematics Cut Score Proportions Chart, Round 3		
	NAEP-like Score	Percent at or Above
	510	0.0
	505	0.0
	500	0.0
	495	0.0
	490	0.0
	485	0.0
	480	0.0
	475	0.0
	470	0.0
	465	0.0
	460	0.0
	455	0.0
	450	0.0
	445	0.0
	440	0.0
	435	0.0
	430	1.0
	425	2.0
	420	3.0
	415	5.0
	410	7.0
	405	11.0
	400	14.0
	395	19.0
	390	24.0
	385	29.0
Cut Score	380	34.0
	375	40.0
	370	46.0
	365	52.0
	360	57.0
	355	62.0
	350	67.0
	345	72.0
	340	76.0
	335	80.0
	330	83.0
	325	86.0
	320	88.0
	315	91.0
	310	92.0
	305	94.0
	300	95.0
	295	96.0
	290	97.0
	285	97.0
	280	98.0
	275	98.0
	270	99.0
	265	99.0
	260	99.0
	255	99.0
	250	99.0
	245	99.0
	240	99.0
	235	99.0
	230	99.0
	225	99.0
	220	100.0
	215	100.0

Figure 19: Illustration of a Cut Score Proportion Chart identifying the percent of students scoring at or above every fifth score level

Selection of Exemplar Items

Exemplar items are used to illustrate what borderline students know and can do. After completing other round 3 activities, panelists will be asked to provide recommendations concerning the selection of exemplar items for their subject. Panelists will be instructed to discuss potential exemplar items with their table group and then provide independent ratings on the basis of whether the mathematics or reading content required by the item seems appropriately matched to the knowledge and skills of the borderline student. Each item will be rated as *Very Good*, *OK*, or *Do Not Use* as an exemplar. Potential exemplar items will be those mapping between the cut score and 50 points above the cut score.

Process Evaluation

Procedural validity is a necessary, but not sufficient, condition for the validity of the standard setting outcomes. Procedural validity is provided in the form of evidence that the procedures were carried out as intended, and were understood by the panelists. At the end of each round and each day, panelists will be provided with an evaluation form designed to assess their understanding of instructions, tasks, and materials. Five questionnaires are recommended for administration over the course of the panel meetings. Most responses are collected on Likert scales, but several responses are narratives that address specific aspects of the process. These evaluations will be reviewed at the end of each day and any sources of confusion, dissatisfaction, or other concerns will be identified for clarification with individual panelists or the panel as a whole.

In order to allow for comparison of procedural data from NAEP achievement level-setting (ALS) meetings, an effort will be made to keep the evaluation questions largely the same as questions used to evaluate NAEP ALS methods in the past. However, some differences are necessary due to differences between the preparedness and ALS standard-setting procedures. The questionnaires for the different preparedness studies should be the same, to the extent possible. Strong support for procedural validity is demonstrated by consistent mean (average) responses on most items at or above 4.0 on a 1-5 scale.

Initial drafts of the evaluation questionnaires for setting standards on grade 12 mathematics for student preparedness for college and the workplace are presented in Appendix D. The evaluation questionnaires used for reading would be identical, except for the labeling.

In addition to process evaluation questionnaires, a debriefing session will be held at the end of each study. During this session, panelists will be asked about their overall impression of the process, their satisfaction with their borderline student description, what worked best, what did not work so well, and how the process might be improved. Questions about the use of the 0.67 RP criterion might also be asked during this session. These recommendations, along with process evaluation data and input from observers, will be used to evaluate and improve the standard-setting process.

Information Processing

NAEP materials and Governing Board information must be handled according to the strictest security measures. Consequently, vendors must be committed to the strict safeguarding and confidential handling and processing of all items, data, analyses, and reports for each standard-setting study conducted for the Governing Board. Reliable procedures must be implemented to

assure security of materials, and computer program features and all algorithms employed in standard setting must be subject to quality control checks.

The technical and logistic support needed for the modified bookmark procedure is relatively modest. For a given cut score, only one numerical rating is provided by each panelist. The ratings will be on a linear transformation of the NAEP composite scale. Each replicate panel will use a different NAEP-like scale to minimize cross panel influence of the cut scores. This transformation from the NAEP scale is needed so that the panelists are not influenced by the scale values for the current achievement levels, which are publically available. No IRT analysis software will be needed to map panelists' item ratings to the NAEP scale. The key entry of data, computational demands, and overall potential for human error in the processing of the data will be minimal.

The contractor is encouraged to use PCs during the standard setting process, to the extent that this increases efficiency and effectiveness. Procedures to ensure the security of these PCs during the meeting must be in place. All secure material distributed to panelists must be identified as such. Each facilitator will need to account for all materials distributed each day. The counts are to be verified when the materials are "checked in" at the staff office. Secure materials including PCs must be locked up at all times, or be under direct watch of project staff. Only security staff of the meeting site facility may have access to the storage space for secure materials.

Quantitative data analyses for each workshop should include, but not be limited to, the following:

- descriptive statistics on the recruited panelists and those who participated in the study;
- median cut score rating for each round for each replicate panel and statistical comparisons of the cut scores for the two panels in each subject;
- standard error of the panel cut score for each round;
- measures of interjudge agreement (rater locations) within and across replicate panels;
- frequency distributions and summary statistics of panelists' responses to all evaluation questionnaires; and
- evidence of procedural validity.

Additionally, analyses will contain a record of the qualifications of participants, recorded written comments, and anecdotal evidence related to the perceptions of participants about the adequacy of the materials, procedures, and results.

This design features replicate panels, and the analysis of data and interpretation of results must focus on this feature.

A system of secure file transfer will be instituted for exchange of data between the vendor and NAEP Alliance Contractors such as ETS, Fulcrum, Pearson, and Westat.

Validity Evidence

Throughout the standard-setting process, procedures to evaluate the efficacy of the process and to collect evidence that the performance standards are reasonable and appropriate must be implemented (Hambleton & Pitoniak, 2006). Perhaps the most important question in the setting of educational performance standards today is whether standard-setting panelists comprehend their tasks and are able to render the types of judgments asked of them (McGinty, 2005). For purposes of these standard-setting studies, evidence on two types of validity must be collected: procedural and internal (Table 1). As shown in Table 1, panelists' understanding of the process is generally addressed by both procedural and internal validity data. The reasonableness of the cut scores is generally addressed by external validity data, which is not covered under these particular standard-setting studies. However, the Governing Board is collecting data that could inform the external validity of the cut scores as part of the overall 12th Grade NAEP Preparedness Research Program, and the Governing Board may choose to have those data used in the judgmental standard setting studies.

Table 1: Types of validity

Type	Definition	Evidence
Procedural	Procedures are reasonable, were carried out as intended, and were understood by panelists	- Explicit documentation of the design and procedures - Panelist feedback on the design and process - Documentation of the process as implemented including any alternations from the original design so as to allow for replication
Internal	Methods were consistent and ratings indicated increasing internal consistency across rounds and panelists	- Comparisons of cut scores using the same method on separate occasions - Variability of each panelists' cut scores across rounds - Variability of cut scores among panelists - Variability of cut scores across groups (item pool sets)

Procedural Validity

Procedural validity generally means that procedures are reasonable, were carried out as intended, and were understood by panelists. Evidence for procedural validity consists largely of documenting that the intended procedures were correctly performed and assessing panelists' understanding through direct questions.

The vendor will be expected to make extensive use of process evaluation questionnaires to document through direct questioning that procedures were implemented well and that panelists, consequently, understood their tasks. Process questionnaires are to be administered at the conclusion of training, each day, each round, and at the end of the meeting for each cut score study. Many items on the questionnaires have a history of use in previous NAEP standard-setting projects and responses will be compared for the method used in the preparedness cut score setting and the methods used in previous achievement levels-setting meetings, although caution will need to be exercised since the purpose of ALS standard setting and preparedness standard setting are different. Most questions will be identical to those administered in the past to allow for comparison. Examples of process evaluation questionnaires are contained in Appendix D. These are comprised of questions with Likert-response scales, but several questions are open-ended to allow for more detailed responses on specific aspects of the process.

At the conclusion of each meeting, a general evaluation questionnaire will be administered. This will ask panelists for their thoughts on the entire process and outcomes (see Table 2). It will also include questions about the adequacy of the amount of time allocated to each task on a scale from 1=far too short to 5=far too long.

Table 2: General evaluation questions used at the completion of each meeting

Question
The most accurate description of my level of <i>confidence</i> in the cut score recommendation I provided was... (5=Totally confident 1=Not at all confident)
I would describe the <i>effectiveness</i> of this standard-setting method as... (5=Highly effective, 1=Not at all effective)
I feel that this standard-setting process provided me an opportunity to use my <i>best judgment</i> to recommend a cut score (5=To a great extent, 1=Not at all)
The <i>instructions</i> on what I was to do during each round were... (5=Absolutely clear, 1=Not at all clear)
My <i>understanding</i> of the tasks I was to accomplish during each round... (5=Totally adequate, 1=Totally inadequate)
The <i>amount of time</i> I had to complete the tasks I was to accomplish during each round was: (5=Far too long, 1=Far too short)

At the completion of training, panelists are also to be asked to respond to questions about their understanding of and the adequacy of time allotted for discussion of the advance briefing materials, method, response probability, the mathematics or reading framework, and specific attributes of the method (e.g., item map, Ordered Item Book, response probability).

Following each round, panelists are to be asked for their feedback on the amount of time allotted to each activity in that round; the clarity of instructions by task; the clarity of the topic presentation (if relevant); their understanding of any new concepts presented, tasks, feedback, the RP criterion, and the borderline student description. They should also be asked to indicate the degree to which they felt pressured to recommend cut scores close to those recommended by other panelists.

Internal Validity

Internal validity generally refers to indirect, but more objective evidence that panelists understand their tasks and are capable of performing the tasks expected of them. Internal validity data may be quite method-specific. In the modified bookmark procedure, data should be collected pertaining to the relative frequency of bookmarked items by item type and content, the exercise of independent judgment, panelists' understanding of the RP criterion, and so forth. Internal validity will include comparison of cut scores across the two Ordered Item Books, which differ only in the items presented to the panelists in each replicate panel. Since these items have been carefully tailored to be essentially equivalent across item type and difficulty level, cut scores for the two replicate panels can be used to evaluate the variability inherent in the process being used. Additionally, variability of cut scores within an Ordered Item Book and within a table also provide a valuable measure of the variability of the cut scores, and so, of the internal validity of the procedure.

The interpretation of internal validity data might also be method specific. Trends in cut scores and trends in the variability of cut scores across rounds may have different meanings depending on the

method. With methods that allow panelists to more or less directly recommend cut scores, such as the modified bookmark, some variability in cut scores may be expected because it indicates that panelists are exercising independent judgment and are responding to the feedback (Hambleton & Pitoniak, 2006). Both intra- and inter-rater consistency will be assessed. Normality of cut scores within rounds and levels can be assessed through the Shapiro-Wilk (1965) test of normality performed on the cut scores within a round. The number of panelists who increased, decreased, or did not change their cut score recommendation should also be assessed by round. In the modified bookmark method, it is expected that the panelists' cut scores will vary somewhat across rounds with the greatest variance between rounds 1 and 2. This will be due to their responsiveness to the feedback provided. The average absolute difference (AAD) of cut scores from the median by round should be also calculated for each round to indicate variability amongst panelists. It is expected that the differences among panelists' cut score recommendations will get smaller over rounds and that the most convergence will occur between rounds 1 and 2, as has been found in previous studies.

Finally, the Governing Board's stipulation that the method selected for the standard-setting meeting be implemented the same way it was in the pilot study (when there is a pilot study) means that the reliability of the method can be assessed through replication both for panels within a workshop and across pilot and operational assessments. Reliability is a key factor in assessing validity. To this end, cut scores will be compared for the replicate panels within the pilot study and operational study and across panels for the pilot and operational studies. Results from this process allows analysis of the statistical equivalence of the final cut scores when the process is implemented by two replicate groups on a single occasion and by two equivalent groups on different occasions. This design provides a rare opportunity for collecting such important information on the reliability of the results across studies.

Reporting

Documentation of the entire process and clear, comprehensive reporting of the procedures and results are mandatory. This design document provides the detailed steps to be implemented in order to maximize standardization and comparability of procedures across studies. Similarly, the reports for the studies must be standardized in order to maximize the comparability of information across studies and the utility of each study to the overall NAEP preparedness research effort. In the event that there are multiple contractors for the standard-setting workshops, the Governing Board will facilitate communication among the contractors to ensure sufficient comparability in salient elements of the reports. This may involve conference calls and face-to-face meetings with the contractors, which will be called for and arranged by the Governing Board.

An example of a report on a process using a modified bookmark methodology is found at www.nagb.org/publications/2006-g12th-econ-process-report.pdf. There will be a process report on the pilot study workshop and a process report on the operational workshop. These two reports are to focus on the replicate panels such that the reports document information for each panel and provide interpretative information about the comparison of results across the replicate panels within each subject group. For higher education, there will be a report for the pilot study workshop, including both mathematics and reading panels, as well as for the operational workshop, including both mathematics and reading panels. For occupational job training reports, there will be a report for the pilot study and for each occupational workshop, including both mathematics and reading panels.

Finally, there will be a technical report to document the technical decisions that guided the process and to document the computational procedures used in the studies. An example of a technical report for a standard setting study using a modified bookmark methodology is found at www.nagb.org/publications/2006-g12th-econ-tech-report.pdf. A single technical report will be submitted to provide complete documentation for both the pilot study workshop and operational workshop. For higher education, there will be a single technical report. For each occupational job training area, there will be a single technical report.

References

- ACT. (2010). *Developing achievement levels on the 2009 National Assessment of Educational Progress in science for grades four, eight, and twelve: Process report*. Iowa City, IA: Author.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 433-470). Westport, CT: American Council on Education/Praeger.
- McGinty, D. (2005). Illuminating the "black box" of standard setting: An exploratory qualitative study. *Applied Measurement in Education*, 18, 269-287.
- Schulz, E. M., Lee, W., & Mullen, K. (2005). A domain-level approach to describing growth in achievement. *Journal of Educational Measurement*, 42, 1-26.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika*, 52, 591-611.

Appendix **A**

Sample Agenda for Judgmental Standard Setting

Note that this example is for one subject of a workshop. An agenda will be required for each subject in each workshop.



AGENDA

SETTING STANDARDS ON THE 2009 GRADE 12 MATHEMATICS NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP) FOR 12TH GRADE STUDENT PREPAREDNESS IN MATHEMATICS FOR PLACEMENT IN A CREDIT-BEARING COLLEGE COURSE

Dates
Location

Wednesday, Date

Camino Real

5:00 - 8:00 PM

Meeting Registration

Please stop by our registration table located in Camino Real to turn in completed security and other forms and to receive your name badge. If you do not arrive at the hotel in time to register Wednesday night, please visit our registration table outside the Camino Real at 8:00 Thursday morning.

Get-Acquainted Social Time

We hope that you will stay around after you have registered to get to know the other meeting participants as well as the project staff.

Staff will be available between 5:00-8:00pm to get you registered and to help you get acquainted with your temporary surroundings.

Thursday, Date

Lobby area outside Camino Real

8:00 AM Registration

Camino Real (Panelists from Entire Workshop)

8:30 AM Welcome and Introductions, *Project Director*

8:45 AM General Orientation to the NAEP, COR, *The National Assessment Governing Board*

9:10 AM Orientation to the Standard-setting Method, *Process Facilitator*

10:30 AM Break

11:15 AM Taking a NAEP Exam, *Project Director and Staff*

11:45 AM Scoring the NAEP Exam, *Project Director and Staff*

Olivares (lower level)

12:15 PM LUNCH

Camino Real One (Mathematics Group A and Group B Panelists)

1:00 PM The NAEP Mathematics Framework, *Content Facilitator*

1:45 PM Review, Discussion, and Modification of Draft Descriptions of Borderline Performance, *Content Facilitator*

2:45 PM Break

Camino Real One (Group A) and Camino Real Two (Group B)

3:00 PM Panel Review of Common Constructed Response Items, *Facilitators*

Evaluation #1

6:00 PM Adjourn

Friday, Date

Camino Real One (Group A) and Camino Real Two (Group B)

8:00 AM Table Group Review of Remaining Constructed Response Items, *Facilitators*

10:00 AM Break

10:15AM Independent Review of Ordered Item Book (take breaks as needed), *Process Facilitator*

Olivares

12:00 PM LUNCH

Camino Real One (Group A) and Camino Real Two (Group B)

1:00 PM Table Group Discussion of Ordered Item Book, *Process Facilitator*

2:30 PM Break

Camino Real One (Group A and Group B)

2:45PM Reaching a Common Understanding of Borderline Performance Descriptions),
Content Facilitator
(Descriptions will have been edited by content facilitator and prepared for panelists' discussion and additional modification prior to Round 1 bookmark placements)

3:45PM Break

Camino Real One (Group A) and Camino Real Two (Group B)

4:00 PM Round 1 Bookmark, *Process Facilitator*
(Note: Revised borderline performance descriptions will be prepared for distribution to panelists for use in Round 1 bookmark placement process)
Evaluation #2

5:30 PM Adjourn

Saturday, Date

Camino Real One (Group A) and Camino Real Two (Group B)

- 8:00 AM Overview of Day's Activities and Feedback from Round 1, *Process Facilitator*
- Review Activities for Today
 - Feedback from Round 1 and Instructions on Use
 - Panel Cut Score
 - Cut Score Distribution Chart
 - Booklet Score Plots
 - Booklet Score Charts
 - Item Score Tables
 - Review and Discussion of Borderline Booklets for Common Form

9:45 AM Break

- 10:00 AM Independent Review of Student Booklets, *Process Facilitator*
- Background
 - Booklet Item Maps

11:00 AM Table Group Discussion of Student Booklets, *Process Facilitator*

Olivares

11:30 AM LUNCH

Camino Real One (Group A and Group B)

- 12:30 PM Reaching a Common Understanding of Borderline Performance Descriptions),
Content Facilitator
(Descriptions will have been edited by content facilitator and prepared for panelists' discussion and final modification prior to Round 2 bookmark placements)

Camino Real One (Group A) and Camino Real Two (Group B)

1:30 PM Round 2 Cut Score Recommendation, *Process Facilitator*

2:00 PM Evaluation #3 and Break

- 4:00 PM Feedback from Round 2, *Process Facilitator*
- Feedback from Round 2
 - Cut Scores
 - Cut Score Distribution Chart
 - Booklet Feedback
 - Impact Data
 - Discussion of Feedback Data
 -

4:45 PM Round 3 Cut Score Recommendation, *Facilitator*
Evaluation #4

5:15 PM Adjourn

Sunday, Date

Camino Real One (Group A) and Camino Real Two (Group B)

8:30 AM Overview of Day's Activities and Feedback from Round 3, *Process Facilitator*

- Feedback from Round 3
 - Cut Scores
 - Cut Score Distribution Chart
 - Impact Data and Discussion

Camino Real One (Group A and Group B)

9:15 AM Review Impact Data Across Groups A and B and Discuss
Impact Data Questionnaire, *Process Facilitator*

10:15 AM Evaluation #5 and Break
(Panelist may bring luggage down for check out.)

Camino Real One (Group A) and Camino Real Two (Group B)

10:45 AM Debriefing, *Process Facilitator*

Camino Real (Panelists from Entire Workshop)

11:30 AM Wrap-Up, *Project Director and COR*

12:00 PM Adjourn

Appendix **B**

Definition of 12th Grade Student Preparedness

Definition of 12th Grade Student Preparedness Resolution

March 2009

The National Assessment Governing Board recognizes that it is necessary to have a clear definition of the term “12th grade student preparedness” for use in explaining student achievement results in National Assessment of Educational Progress reports. This definition should be understandable and useful to the public and at the same time must be consistent with NAEP’s characteristics and limitations. No single, generally accepted definition of “12th grade student preparedness” currently exists, whether for NAEP or for public policy use more broadly. In order to proceed with the program of 12th grade preparedness research proposed for NAEP, a working definition of 12th grade student preparedness is essential. The working definition will be subject to adjustment and revision based on the outcome of the planned program of research. The working definition is conditioned on the following three considerations:

1. Preparedness is not intended to represent success in postsecondary education and training.
2. Preparedness in the NAEP context must be limited to academic qualifications for postsecondary education and workplace training.
3. Preparedness for workplace training is intended to have the same meaning for selected occupations in both the military and civilian sectors and is based on the assumption that similar occupations in both the military and civilian sectors require approximately equal reading and mathematics knowledge and skills to qualify for entry.

With these considerations in mind, the Governing Board affirms the following, derived from recommendations of the 12th Grade Technical Panel, as the working definition of “12th grade student preparedness” for the purpose of conducting the proposed program of 12th grade preparedness research:

Preparedness for college refers to the reading and mathematics knowledge and skills necessary to qualify for placement into entry level college credit courses that meet the general education requirements without the need for remedial coursework in mathematics or reading.

Preparedness for workplace refers to the reading and mathematics knowledge and skills needed to qualify for an occupation’s job training program; it does not necessarily mean that the qualifications to be hired for a job have been met.

Appendix

C

Example of Impact Data Questionnaire

2009 NAEP Standard Setting for (insert occupation/field)

QUESTIONNAIRE ON IMPACT DATA

You, together with the other panelists, have set a cutscore to represent preparedness for placement in a college-level credit-bearing course in mathematics. Borderline performance is defined as:

The mathematics knowledge and skills needed to qualify for placement into entry level college credit courses that meet the general education requirements without the need for remedial coursework in mathematics.

The final description of borderline performance for mathematics used in setting your cutscore is

XX
 XX
 XX
 XX
 XX
 XX
 XX

You have been given information on student performance at the cutscore you defined (this is the final cutscore, the median cutscore for your panel at the end of round 3), as well as for students at the Basic, Proficient, and Advanced levels. The percentage at or above the cutscore is the percentage of all student who took the NAEP assessment that scored at that cutscore or higher.

In this questionnaire you are asked to evaluate the cutscore set by your panel in light of the data provided (i.e., information about the percentages of all NAEP students scoring at or above your cutscore). We are interested in knowing whether or not this information about actual student performance is compelling enough to you that you would alter the cutscore if you had the opportunity to do so.

Please fill in the numbers in the following sentences:

Your panel set the final cutscore at _____. This means that approximately _____ percent of students in grade 12 would score at or above the cutscore for preparedness in mathematics.

Please mark the boxes below that correspond to the statements that best characterize your opinions regarding this percentage and the cutscore your panel set.

1. Given your understanding of borderline student performance does this percentage reflect your expectations about the proportion of students whose NAEP score would indicate at least minimal preparedness for placement in a college-level mathematics course?
☐ Yes (Please skip to Number 4.)
☐ No (Please continue to Number 2.)
2. Having seen the data on the percentage of students whose score on the NAEP was at or above the cutscore your panel set, would you change the cutscore set by your panel to recommend to the Governing Board if you could?
☐ Yes (Please continue to Number 3.)
☐ No (Please skip to Number 4.)
3. Please mark the box corresponding to the response that indicates *how you would change the final cutscore*. Changing the final cutscore would make these percentages more in line with your expectations about the proportions of students taking the Mathematics NAEP whose score would indicate preparedness for a college-level, credit-bearing mathematics course. *You must give a cutscore if you recommend a change.*
☐ Make no change. I am satisfied with the cutscore.
☐ Raise the cutscore to represent that a *smaller* percentage of students is prepared for placement in a college level mathematics course. I want to *raise* the cutscore to _____.
☐ Lower the cutscore to represent that a *larger* percentage of students is prepared for placement in a college level mathematics course. I want to *lower* the cutscore to _____.
4. What recommendations do you wish to make to the National Assessment Governing Board regarding the cutscore set for college level mathematics course placement?
5.
☐ I recommend that the cutscore be reported as set.
☐ I recommend changes consistent with my answers above. If you wish, comment on the magnitude of change you would recommend.

Appendix

D

Examples of Process Evaluation Questionnaires

2009 NAEP Mathematics Judgmental Standard Setting for College-Level Course Placement in Mathematics

Date

Process Evaluation Questionnaire No. 1

Please take a few minutes to complete this Process Evaluation Questionnaire so that the procedures used in this study can be evaluated. Your evaluation is a key element in the design of the process. Your panelist identification number is used for analysis purposes only. Your responses to this questionnaire will be held in strict confidence and will be analyzed only in conjunction with those of other panelists who participated in this meeting and other meetings of the 2009 NAEP research on academic preparedness of 12th grade students for entry-level credit bearing college coursework.

SECTION I: Advance Materials

If you did not receive any advance materials prior to this meeting, check here ☐ and skip to Section II of this questionnaire.

1. The advance materials I received were adequate to prepare me to fulfill my role in this meeting:	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
2. The organization of the advance materials I received for this meeting was:	Very Good ☐	☐	Acceptable ☐	☐	Very Poor ☐

SECTION II: General Orientation to NAEP Program

3. The amount of time allocated for the General Orientation to the NAEP Program was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
4. The explanation of the NAEP in general was:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
5. The explanation of the development of the NAEP Mathematics was:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
6. I understand the purpose of this NAEP standard setting workshop.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION III: General Introduction to Standard Setting Method

7. The amount of time allocated for the General Introduction to the NAEP standard setting process was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
8. I believe my perspectives and experiences will be important in the NAEP standard setting process.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
9. I understand the difference between criterion-referenced and norm-referenced standards.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION IV: Taking the NAEP Exam

10. Taking the NAEP Mathematics was an informative experience.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
11. Taking the NAEP Mathematics gave me a good idea of what is expected of students.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION V: Orientation to the Bookmark-based method

12. The amount of time allocated for the orientation to the methodology was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
13. The overview of the method to be followed in this meeting was:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
14. The explanation of how an item map is constructed was:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
15. I think I will be comfortable using a 2/3 or 0.67 probability to interpret the location of an item on my map.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
16. The explanation of the information in my Ordered Item Book (OIB) was:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐

SECTION VI: Mathematics Framework

17. The amount of time allocated for the Mathematics Framework presentation was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
18. The presentation of the Mathematics Framework was:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
19. The presentation of the Mathematics Framework had about the right level of detail.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION VII: Constructed Response Item Review

20. The amount of time allocated for the group review of items was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
21. The instructions on what I was to do in the Item Review were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
22. My understanding of our tasks in the Item Review was:	Totally Adequate ☐	☐	Somewhat Adequate ☐	☐	Totally Inadequate ☐
23. The group work on the common constructed response items was:	Very Useful ☐	☐	Somewhat Useful ☐	☐	Not at all Useful ☐

24. Were there any questions or concerns that were NOT answered or addressed in advance of your coming here?
Please indicate those here.

25. Of the advance materials you received, what was most helpful?

26. Please use the space below to provide additional comments concerning the adequacy, appropriateness, usefulness, or organization of the materials you received prior to this meeting.

27. Please identify the most helpful information you received or the most useful activity in which you participated today.

28. Please comment on areas of strength and areas for improvement in the Mathematics Framework training session.

29. Please comment on areas of strength and areas for improvement on the constructed response item review and the use of mathematics knowledge and skills.

Additional Comments

30. Please use the space below to provide any additional comments or suggestions concerning the portions of the standard setting process you have experienced to this point.

Thank You!

Your responses will help to improve the process of setting standards.

2009 NAEP Mathematics Judgmental Standard Setting for College-Level Course Placement in Mathematics

Date

Process Evaluation Questionnaire No. 2

Please take a few minutes to complete this Process Evaluation Questionnaire so that the procedures used in this study can be evaluated. Your evaluation is a key element in the design of the process. Your panelist identification number is used for analysis purposes only. Your responses to this questionnaire will be held in strict confidence and will be analyzed only in conjunction with those of other panelists who participated in this meeting and other meetings of the 2009 NAEP research on academic preparedness of 12th grade students for entry-level credit bearing college coursework.

SECTION I: Remaining Constructed Response Item Review

1. The amount of time allocated for the table group Item Review was:	Far Too Long ☐	About Right ☐	Far Too Short ☐
2. The table group review of the remaining constructed response items was:	Very Useful ☐	Somewhat Useful ☐	Not at all Useful ☐
3. I understand the score levels of constructed response items.	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐

SECTION II: Independent Review of Ordered Item Book (OIB)

4. The amount of time allocated for the independent OIB review was:	Far Too Long ☐	About Right ☐	Far Too Short ☐
5. The instructions on what I was to do for the independent OIB review were:	Absolutely Clear ☐	Somewhat Clear ☐	Not at All Clear ☐
6. I understand how to use my item map with the OIB.	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐
7. I was comfortable working through the OIB on my own.	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐
8. The ordering of the items in the OIB agreed with my perceptions of the relative difficulty of the items.	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐
9. The work to identify mathematics knowledge and skills associated with items in the OIB helped me understand what can make one item harder than others.	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐

SECTION III: Table Discussion of Ordered Item Book (OIB)

10. The amount of time allocated for the table discussion of the OIB was:	Far Too Long ☐	About Right ☐	Far Too Short ☐
---	--	---	---

11. The instructions on what we were to do in the table discussion of the OIB were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
12. The table discussion of the OIB was:	Very Useful ☐	☐	Somewhat Useful ☐	☐	Not at all Useful ☐
13. I feel I made a valuable contribution to my table group's discussion.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
14. I feel my perspective is being heard by others in my table group.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
15. I feel that I was being pressured to agree with others in my table group.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION IV: Describing the Borderline Student

16. The amount of time allocate for developing the description was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
17. My understanding of the NAEP Mathematics was sufficient for the task.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
18. The panel participation in developing the description was:	Very Useful ☐	☐	Somewhat Useful ☐	☐	Not at All Useful ☐
19. The description of borderline performance is a reasonably complete and a comprehensive statement of what a student should know and be able to do for placement in a college level credit-bearing mathematics course.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
20. My own level of satisfaction with the borderline performance description is:	Very Satisfied ☐	☐	Somewhat Satisfied ☐	☐	Not at All Satisfied ☐
21. I feel comfortable about my understanding of how the mathematics knowledge and skills relate to the borderline performance description.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
22. I feel comfortable using the borderline performance description to develop the idea of a minimally qualified student.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION V: Round 1 – Bookmark Placement

23. At the time I provided the round 1 bookmark placement, my understanding of borderline performance was:	Totally Adequate ☐	☐	Somewhat Adequate ☐	☐	Totally Inadequate ☐
24. I was comfortable using the description of borderline performance.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
25. I believe my round 1 bookmark placement is consistent with the description of the level of preparedness required for placement in a college-level course in mathematics..	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

26. The amount of time allocated for placing the bookmark was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
27. The instructions on how I was to place my bookmark were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
28. My understanding of how to use the borderline performance description to choose my bookmark was:	Totally Adequate ☐	☐	Somewhat Adequate ☐	☐	Totally Inadequate ☐
29. The most accurate description of my level of confidence in my round 1 bookmark placement is:	Totally Confident ☐	☐	Somewhat Confident ☐	☐	Not at All Confident ☐
31. The mathematics knowledge and skills required by the items around my bookmark seem consistent with those required in the borderline performance description.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
32. When choosing your cut score (bookmark), how difficult was it to take into account how the mathematics knowledge and skills relate to the borderline performance description.	Not at All Difficult ☐	☐	Somewhat Difficult ☐	☐	Very Difficult ☐
34. I feel comfortable using a 2/3 or 0.67 probability for defining mastery in order to place the bookmark.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
35. When choosing your cut score (bookmark), how difficult is it to think about and use the 2/3 or 0.67 criterion?	Not at All Difficult ☐	☐	Somewhat Difficult ☐	☐	Very Difficult ☐

36. Please comment on the areas of strength and areas for improvement in the OIB review.

37. Please comment on the areas of strength and areas for improvement in the development of the description of borderline performance .

Please use the space below to provide additional comments concerning the round 1 bookmark placement. (Suggestions for improvement or amplification of instructions or procedures would be particularly helpful):

38. Please comment on areas of strength and areas for improvement in round 1 bookmark placement:

39. If you experienced any particular difficulties, please identify those here.

Additional Comments

40. Please use the space below to provide any additional comments or suggestions concerning the portions of the standard setting procedure you have experienced to this point:

Thank You!

Your responses will help to improve the process of setting standards.

2009 NAEP Mathematics Standard Setting for College-Level Course Placement in Mathematics

Date

Process Evaluation Questionnaire No. 3

Please take a few minutes to complete this Process Evaluation Questionnaire so that the procedures used in this study can be evaluated. Your evaluation is a key element in the design of the process. Your panelist identification number is used for analysis purposes only. Your responses to this questionnaire will be held in strict confidence and will be analyzed only in conjunction with those of other panelists who participated in this meeting and other meetings of the 2009 NAEP research on academic preparedness of 12th grade students for entry-level credit bearing college coursework.

SECTION I: Feedback from Round 1

1. I understand how the round 1 median cut score was computed.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
2. I understand what students at the round 1 median cut score can do.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
3. I understand the Rater Location Feedback (where my round 1 cut score was in comparison to the round 1 median cut score).	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
4. I understand the cut score dispersion chart (bar graph of cut scores).	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION II: Reviewing the Borderline Performance Description

5. The amount of time allocate for reviewing the description was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
6. The instructions for reviewing and discussing the borderline performance description were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
7. The discussion of the description after conducting round 1 helped me understand the borderline performance and improve the description.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
8. The whole group participation in discussing the borderline performance description and revising the description was:	Very Useful ☐	☐	Somewhat Useful ☐	☐	Not at All Useful ☐

SECTION III: Whole Booklet Feedback Tasks

9. The amount of time allocated for the borderline booklet exercise was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
10. The instructions I received for the borderline booklet exercise were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐

11. The purpose of the borderline booklet exercise was:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
12. The borderline booklet exercise helped me understand how student booklets illustrate performance at a given cut score.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
13. The borderline booklet exercise helped me understand that student performance on individual items may vary even at the same cut score.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
14. The amount of time allocated for the table group whole booklet review was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
15. The instructions I received for the table group whole booklet review were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
16. The purpose of the table group whole booklet review was:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
17. The item maps showing the items in the booklets were useful.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
18. The item score tables were useful.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
19. The booklet score <u>chart</u> was useful.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
20. The booklet score <u>plots</u> were useful.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
21. I understand the information presented in the booklet score <u>chart</u> .	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
22. I understand the information presented in the booklet score <u>plot</u> .	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
23. I understand the information in the item score tables.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION IV: Round 2 Cut Score Recommendation					
24. At the time I provided the round 2 cut score recommendations, my understanding of the borderline performance description was:	Totally Adequate ☐	☐	Somewhat Adequate ☐	☐	Totally Inadequate ☐
25. I believe my round 2 cut score recommendation was consistent with the borderline performance descriptions.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
26. The amount of time allocated for my round 2 cut score recommendation was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐

27. The instructions I received for recommending the round 2 cut score were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
28. My level of understanding of how I was to choose a cut score for round 2 was:	Totally Adequate ☐	☐	Somewhat Adequate ☐	☐	Totally Inadequate ☐
29. The most accurate description of my level of confidence in my round 2 cut score recommendation is:	Totally Confident ☐	☐	Somewhat Confident ☐	☐	Not at All Confident ☐
30. I felt pressure to recommend a cut score that is close to those recommended by other panelists.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
31. I was comfortable choosing scale values instead of placing a bookmark to recommend a cut score.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
32. The work with the whole booklets was helpful for setting my round 2 cut score.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
33. The booklet score chart was helpful to me for selecting a cut score.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
34. I was comfortable locating my cut score selection in both the OIB and the booklet score chart.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
35. As I was choosing a cut score, I felt comfortable about how the mathematics knowledge and skills of the items near the cut score related to the borderline performance description.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
36. When choosing your cut score (bookmark), how difficult is it to take into account how the mathematics knowledge and skills related to the borderline performance description?	Not at All Difficult ☐	☐	Somewhat Difficult ☐	☐	Very Difficult ☐
37. I feel comfortable using a 2/3 or 0.67 probability as defining mastery in order to place the bookmark.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
38. When choosing your cut score (bookmark), how difficult is it to think about and use the 2/3 or 0.67 criterion?	Not at All Difficult ☐	☐	Somewhat Difficult ☐	☐	Very Difficult ☐

Additional Comments

39. Please use the space below to provide additional comments concerning the clarity and completeness of the instructions you received, the adequacy of the time available, your level of understanding and confidence, or any other aspects of the feedback provided before the second round.

40. Please comment on any particular difficulties you experienced in making your round 2 cut score selection. Do you have suggestions for improvement?

41. Please use the space below to provide additional comments or suggestions concerning the portions of the standard setting procedure you have experienced to this point.

Thank You!

Your responses will help to improve the process of setting standards.

2009 NAEP Mathematics Standard Setting for College-Level Course Placement in Mathematics

Date

Process Evaluation Questionnaire No. 4

Please take a few minutes to complete this Process Evaluation Questionnaire so that the procedures used in this study can be evaluated. Your evaluation is a key element in the design of the process. Your panelist identification number is used for analysis purposes only. Your responses to this questionnaire will be held in strict confidence and will be analyzed only in conjunction with those of other panelists who participated in this meeting and other meetings of the 2009 NAEP research on academic preparedness of 12th grade students for entry-level credit bearing college coursework.

SECTION I: Feedback from Round 2 Cut Score Recommendation

1. I understand how the round 1 median cut score was computed.	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐
2. I understand what students at the round 2 median cut score can do.	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐
3. I understand the Rater Location Feedback (where my round 2 cut score was in comparison to the round 2 median cut score).	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐
4. I understand the cut score dispersion chart (bar chart).	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐
5. I understand the feedback on the booklet score <u>chart</u> .	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐
6. I understand the feedback on the booklet score <u>plots</u> .	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐

SECTION II: Impact Data and Discussions

7. I understand the impact data.	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐
8. The instructions I received for using impact data during round 3 were:	Absolutely Clear ☐	Somewhat Clear ☐	Not at All Clear ☐
9. The amount of time allocated for discussing the impact data was:	Far Too Long ☐	About Right ☐	Far Too Short ☐
10. The most accurate description of my level of confidence in using the impact data to recommend cut scores in round 3 is:	Totally Confident ☐	Somewhat Confident ☐	Not at All Confident ☐
11. I feel comfortable about using the impact data to evaluate the reasonableness of my cut score.	Totally Agree ☐	Somewhat Agree ☐	Totally Disagree ☐

SECTION III: Round 3 Cut Score Recommendation

12. At the time I provided the round 3 cut score recommendations, my understanding of the borderline performance description was:	Totally Adequate ☐	☐	Somewhat Adequate ☐	☐	Totally Inadequate ☐
13. I believe my round 3 cut score recommendation is consistent with the borderline performance description.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
14. The instructions I received for recommending the round 3 cut score were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
15. My level of understanding of how I was to choose a cut score for round 3 was:	Totally Adequate ☐	☐	Somewhat Adequate ☐	☐	Totally Inadequate ☐
16. The most accurate description of my level of confidence in my round 3 cut score recommendation is:	Totally Confident ☐	☐	Somewhat Confident ☐	☐	Not at All Confident ☐
17. I felt pressure to recommend a cut score that was close to those recommended by other panelists.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

Additional Comments

18. Please use the space below to provide additional comments concerning the clarity and completeness of the instructions you received, the adequacy of the time available, your level of understanding and confidence, or any other aspects of the third round:

19. Following the third round of the cut score placement, please comment on any particular difficulties you experienced. Do you have suggestions that would improve this situation?

20. Please use the space below to provide additional comments or suggestions concerning the portions of the standard setting procedure you have experienced to this point.

Thank You!

Your responses will help to improve the process of setting standards.

2009 NAEP Mathematics Standard Setting for College-Level Course Placement in Mathematics

Date

Process Evaluation Questionnaire No. 5

Please take a few minutes to complete this Process Evaluation Questionnaire so that the procedures used in this study can be evaluated. Your evaluation is a key element in the design of the process. Your panelist identification number is used for analysis purposes only. Your responses to this questionnaire will be held in strict confidence and will be analyzed only in conjunction with those of other panelists who participated in this meeting and other meetings of the 2009 NAEP research on academic preparedness of 12th grade students for entry-level credit bearing college coursework.

SECTION I: Feedback from Round 3 Cut Score Recommendation

1. I understand the round 3 median cut score.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
2. I understand what students at the round 3 median cut score can do.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION II: Impact Data Questionnaire

3. The amount of time allocated for the impact data questionnaire was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐
4. I understand the round 3 impact data.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
5. The instructions I received for completing the impact data questionnaire were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
6. I understood how to complete the impact data questionnaire.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐

SECTION III: Rounds 1 through 3

7. The instructions on what I was to do during each round were:	Absolutely Clear ☐	☐	Somewhat Clear ☐	☐	Not at All Clear ☐
8. My understanding of the tasks I was to accomplish during each round was:	Totally Adequate ☐	☐	Somewhat Adequate ☐	☐	Totally Inadequate ☐
9. The most accurate description of my level of confidence in the cut score recommendations I provided was:	Totally Confident ☐	☐	Somewhat Confident ☐	☐	Not at All Confident ☐
10. The amount of time I had to complete the tasks I was to accomplish during each round was:	Far Too Long ☐	☐	About Right ☐	☐	Far Too Short ☐

11. I would describe the effectiveness of this standard setting method as:	Highly Effective ☐	☐	Somewhat Effective ☐	☐	Not at All Effective ☐
12. I felt my input was valued and considered by others in my group.	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐
13. I felt pressured by others in my group to make my cut score recommendation agree with theirs.	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐
14. I felt pressured by staff to make cut score recommendation higher or lower.	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐
15. I felt pressured by staff to keep my cut score recommendation the same.	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐

SECTION IV: The Overall NAEP Standard Setting Process

16. I understand the purpose of this meeting.	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
17. I feel that this standard setting process provided me an opportunity to use my best judgment to recommend a cut score to represent preparedness for college course placement on the NAEP Mathematics assessment.	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐
18. I feel that this standard setting process has produced a cut score that is <u>defensible</u> .	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐
19. I feel that this standard setting process has produced cut score that will generally be considered <u>reasonable</u> .	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐
20. I feel that the panel in this meeting is widely inclusive of groups that should have a say in setting NAEP Mathematics cut scores for college-level course placement in mathematics.	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐
21. I feel that the panelists in this meeting are appropriately qualified for setting NAEP Mathematics college-level course placement cut scores.	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐
22. I would be willing to sign a statement (after reading it of course) recommending the use of the cut score resulting from this standard setting process.	☐ Yes, definitely ☐ Yes, probably ☐ No, probably not ☐ No, definitely not				
23. Having observers present influenced my judgments.	To a Great Extent ☐	☐	Somewhat ☐	☐	Not at All ☐
24. During the standard setting process, I found the borderline performance description:	Very Helpful ☐	☐	Somewhat Helpful ☐	☐	Not at All Helpful ☐
25. During the standard setting process, I found the OIB:	Very Helpful ☐	☐	Somewhat Helpful ☐	☐	Not at All Helpful ☐

26. During the standard setting process, I found the Primary Item Map:	Very Helpful ☐	☐	Somewhat Helpful ☐	☐	Not at All Helpful ☐
27. During the standard setting process, I found the Rater Location Data (the location of my cut score relative to the median cut score):	Very Helpful ☐	☐	Somewhat Helpful ☐	☐	Not at All Helpful ☐
28. During the standard setting process, I found the impact data:	Very Helpful ☐	☐	Somewhat Helpful ☐	☐	Not at All Helpful ☐
29. During the standard setting process, I found the Booklet Score Charts:	Very Helpful ☐	☐	Somewhat Helpful ☐	☐	Not at All Helpful ☐
30. During the standard setting process, I found the Booklet Score Plots:	Very Helpful ☐	☐	Somewhat Helpful ☐	☐	Not at All Helpful ☐
31. During the standard setting process, I found the Cut Score Dispersion Chart:	Very Helpful ☐	☐	Somewhat Helpful ☐	☐	Not at All Helpful ☐
32. I would rate the amount of personal attention and assistance I received from the process facilitator (<i>insert facilitator name</i>):	Too Much ☐	☐	About Right ☐	☐	Too Little ☐
33. I would rate the amount of personal attention and assistance I received from the content facilitator (<i>insert facilitator name</i>):	Too Much ☐	☐	About Right ☐	☐	Too Little ☐
34. My employer supported my participation in this meeting:	Totally Agree ☐	☐	Somewhat Agree ☐	☐	Totally Disagree ☐
35. I had to take vacation time in order to attend this meeting:	Totally Agree ☐				Totally Disagree ☐

36. Please evaluate the procedures used to set standards. In particular, please indicate whether you think the procedures you used are useful—do they make sense? Were the descriptions of the procedures and the amount of information and training adequate for you to perform your tasks?

Additional Comments

37. Please comment on the quality of assistance provided by the process facilitator. **In particular**, please indicate whether there are ways in which the process facilitator could have made this a **more positive** experience.

38. Please provide any comments you wish to share regarding the content facilitator. **In particular**, please indicate whether there are ways in which the content facilitator could have made this a **more positive** experience.

39. Please use the space below to provide any additional comments, suggestions, conclusions, or recommendations concerning the overall standard setting process or the borderline performance description that would improve the results from this activity.

40. If you have any comments about the two statements above, or feel another description better summarizes your thought process as you selected your cut score, please write it here.

Thank You!

Your responses will help to improve the process of setting standards.

Appendix **E**

List of Twenty Potential Occupations

List of Potential Exemplar Occupations

The preparedness research studies for placement in job training courses will be for 5-7 occupations to be selected from this list of exemplar occupations.

1. Police patrol officers
2. Nursing aides, orderlies, and attendants
3. Automotive Master mechanics
4. Licensed practical and licensed vocational nurses
5. Preschool teachers, except special education
6. Hairdressers, hairstylists, and cosmetologists
7. Real estate agents
8. Electricians
9. Plumbers
10. Bookkeeping, auditing, accounting clerks
11. Customer service representatives
12. Registered nurses
13. Computer support specialists
14. Civil engineering technicians
15. Electrical engineering technicians
16. Paralegals and legal assistants
17. Medical records and health information technicians
18. Radiologic technologists
19. Dental hygienists
20. Pharmacy technicians